Entropy of convex hulls and Kuelbs-Li inequalities

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

vorgelegt dem Rat der Fakultät für Mathematik und Informatik
der Friedrich-Schiller-Universität Jena

von Diplom-Mathematiker Oliver Kley
geboren am 13.10.1983 in Gotha

Gutachter

1. Prof. Dr. Werner Linde, Jena
2. Prof. Dr. Ingo Steinwart, Stuttgart
3. Prof. Dr. Bernd Carl, Jena

Tag der öffentlichen Verteidigung: 15. Juni 2012

Danksagung

Mein besonderer Dank gilt meinem Betreuer Prof. Dr. Werner Linde. Seine unbedingte Bereitschaft, mir für Diskussionen ein offenes Ohr zu leihen und seine dabei vermittelten Anregungen haben diese Arbeit erst ermöglicht.
Zudem gilt mein Dank Prof. Dr. Mikhail Lifshits für wertvolle Hinweise und Impulse, insbesondere Lemma 30 und Proposition 42 betreffend. Auch sei ihm gedankt für die Erlaubnis, seinen unveröffentlichten Beweis von Proposition 25 einzubinden.
Ich danke Dr. Frank Aurzada zum ersten für die Zusammenarbeit bezüglich Abschnitt 5.1.1 sowie zum zweiten für den Anstoß, die Betrachtungen von Abschnitt 5.1.2 durchzuführen. Drittens war er mir ein hilfreicher Ratgeber bei der Erstellung eines Manuskripts, welches beim Journal of Theoretical Probability eingereicht wurde.
Weiter gilt mein Dank meinem langjährigen Kommilitonen Dr. Johannes Christof, der bei akribischem Lesen einer Vorabversion etliche Lapsus aufdekkte.

Und ich danke Peter nicht lediglich dafür, mein German-English etwas weiter in die Nähe eines English-English gerückt zu haben.

Ich danke meinen Eltern; sie ließen mir jede Freiheit, und sie unterstützten mich in allen meinen Entscheidungen.

Ich danke Lisbeth, Jojo und Thommy, die mich in einer so schwierigen Situation aufgefangen haben – Hit the road, Jack! Und für vier unvergessliche Wochen!

Ich danke allen meinen Freunden. Sie vermochten es, mich im besten Sinne auf andere Gedanken zu bringen.

Ich danke Anne, Eike, Isabelle, Johannes, Linh und Toralf für ein jederzeit bereicherndes Miteinander am Institut.

Schließlich danke ich der Studienstiftung des deutschen Volkes für die finanzielle Unterstützung.

Acknowledgement

I am very grateful to my supervisor Prof. Dr. Werner Linde. His absolute readiness to discuss about the topic and his continual suggestions were main conditions for this work to emerge.
Moreover, I owe gratitude to Prof. Dr. Mikhail Lifshits for valuable hints and impulses, in particular, concerning Lemma 30 and Proposition 42. I am also indebted to him for the permission to include his unpublished proof of Proposition 25.
I thank Dr. Frank Aurzada firstly for collaboration concerning section 5.1.1, secondly for initiating the considerations of section 5.1.2 and thirdly for his worthwile directions in writing a manuscript, which has been submitted to the Journal of Theoretical Probability.
Furthermore I thank my longtime fellow student Dr. Johannes Christof, who read a version of the manuscript carefully and unveiled many lapses.

And I thank Peter not only for moving my German-English a bit nearer to English-English.

I thank my parents. They never urged me in any direction; at the same time, they always supported my decisions.

I thank Lisbeth, Jojo and Thommy who bolstered me when I had quite a hard time – Hit the road, Jack! As well as for four unforgettable weaks!

I thank all my friends. In the best sense, they were able to take my mind off things.

I thank Anne, Eike, Isabelle, Johannes, Linh and Toralf for a steadily enriching cooperation in the institute.

Finally, I thank the German Academic Foundation for financial support.

CONTENTS

1. Einleitung 1

2. Introduction 6

3. Preliminaries 11
3.1 Entropy and covering numbers 11
3.2 Gaussian processes, random variables, and measures 14
3.3 Reproducing kernel Hilbert spaces 17
3.4 Small deviations 20
3.5 Connections between reproducing kernel Hilbert spaces and absolutely convex hulls 23

4. Entropy of convex hulls 29
4.1 A probabilistic approach 29
4.1.1 Kuelbs-Li inequality for different cases 30
4.1.2 Proof of Theorem 2 parts (i) – (iv) 32
4.1.3 Proof of Theorem 2 parts (v) – (vii) 33
4.2 A direct proof of Theorem 2 part (v) 35
4.3 Estimates via majorizing measures in the critical case 37
4.3.1 Basic constructions and properties 38
4.3.2 Proof of Theorem 2, parts (iv)- (vii) 40
4.3.3 Observations on majorizing measures concerning the size of convex hulls 41

5. Critically large entropy 47
5.1 Lower estimates 47
5.1.1 Application: Small deviations of a Gaussian sequence 50
5.1.2 Application: Small deviations of a Volterra type process 57
5.2 Kuelbs-Li inequalities in view of critically large entropy . . . 61
5.3 Variations on the Gao set 67

References 74

1. EINLEITUNG

Das Leitmotiv dieser Arbeit hat eine einfache Gestalt: Um drei Punkte in der Ebene zu überdecken, genügen drei Kreise beliebigen Radius'. Jene drei Punkte bilden die Ecken eines Dreiecks. Um wiederum dieses Dreieck zu überdecken, ist wohl eine größere Anzahl an Kreisen vonnöten. Diese steigt, wenn sich die Kreisradii verkleinern. Das Dreieck ist die konvexe Hülle seiner Eckpunkte; in diesem Sinne können wir etwas allgemeiner die Frage stellen: Wieviele Kugeln eines bestimmten Radius' sind notwendig, um die konvexe Hülle einer Punktmenge in einem linearen normierten Raum zu überdecken, wobei Informationen darüber gegeben sind, wieviele Kugeln benötigt werden, um die Ausgangsmenge zu überdecken. Dieses Problem in all seinen Ausprägungen durchzieht die gesamte Arbeit. Wir präzisieren die Situation:
Sei $[E, d]$ ein metrischer Raum, $T \subseteq E$ eine präkompakte Teilmenge und $B(x_0, \varepsilon) := \{x \in E : d(x_0, x) \leq \varepsilon\}$. Die **Überdeckungszahlen** von T werden definiert als

$$N(T, \varepsilon) := \min\{n \in \mathbb{N} : \exists\, t_1, ..., t_n \in E \text{ so dass } T \subseteq \bigcup_{i=1}^{n} B(t_i, \varepsilon)\}.$$

Auf die Größe $\log N(T, \varepsilon)$ beziehen wir uns als **metrische Entropie** von T. Im folgenden bezeichnet T eine Teilmenge eines Hilbertraumes, welche mit der durch das Skalarprodukt induzierten Metrik ausgestattet ist. Hierbei arbeiten wir nur mit separablen Hilberträumen, selbst wenn dies nicht explizit erwähnt wird.

Die Hauptaufgabe im Sinne obigen Leitmotivs ist es, Schranken für die Überdeckungszahlen der konvexen Hülle einer Menge T zu bestimmen, während uns dabei Informationen über die Menge T gegeben seien. In Carl (1982) und Ball and Pajor (1991) wird die Entropie konvexer Hüllen von Mengen abzählbar vieler Vektoren studiert, gestützt auf Kenntnisse über deren Normen. In Dudley (1987) wurde die Problemstellung derart verallgemeinert, dass lediglich eine Schranke für die Überdeckungszahlen der Ausgangsmenge gegeben ist, ohne spezielles Wissen über die Normen. Es wird

bewiesen, dass $N(T,\varepsilon) \leq c_1\varepsilon^{-\alpha}$ die Ungleichung $\log N(\mathrm{co}(T),\varepsilon) \leq c_2\varepsilon^{-\gamma}$ für $\alpha > 0$ und alle $\gamma > \frac{2\alpha}{2+\alpha}$ impliziert, wobei c_1, c_2 von ε unabhängige Konstanten sind.
Dudleys Untersuchungen waren motiviert von Anwendungen auf dem Gebiet empirischer Prozesse; darüber hinaus findet die Frage nach der Entropie konvexer Hüllen Anwendungen – siehe Carl, Kyrezi, and Pajor (1999) – in verschiedenen Zweigen der Mathematik, etwa in der Approximationstheorie, der Geometrie, der Funktionalanalysis wie auch der Wahrscheinlichkeitstheorie. Nichtsdestoweniger möchten wir unterstreichen, wie natürlich und intuitiv diese Problemstellung ist, was Betrachtungen darüber ungeachtet möglicher Anwendungen interessant macht.
Das oben zitierte Resultat Dudleys wird heutzutage von einer Reihe von Abschätzungen für verschiedene Wachstumsraten von $N(T,\varepsilon)$ begleitet. Ebenso liegen Verallgemeinerungen auf den Fall von Banachräumen des Typs $p \in (1,2]$ vor, vlg. Carl, Kyrezi, and Pajor (1999), Steinwart (2000), Creutzig and Steinwart (2002) und Steinwart (2004).
Anstatt mit der konvexen werden wir mit der **absolutkonvexen Hülle** von T

$$\mathrm{aco}(T) := \{\sum_{i=1}^{n} \lambda_i t_i : \sum_{i=1}^{n} |\lambda_i| = 1,\ t_1,...,t_n \in T,\ n \in \mathbb{N}\}$$

arbeiten, was die Anwendung von Dualitätsargumenten ermöglicht.
Für Funktionen benutzen wir die Notation $f \preceq g$ falls $\limsup_{x\to a} \frac{f(x)}{g(x)} < +\infty$ erfüllt ist mit $a \in [0,+\infty]$. In diesem Fall schreiben wir auch $g \succeq f$. Zudem gebrauchen wir $f \approx g$, wenn $f \preceq g$ und $g \preceq f$. Für Folgen verwenden wir die analoge Notation.
Wir bündeln die Abschätzungen für die wichtigsten Wachstumsraten von $N(T,\varepsilon)$ im nächsten Theorem.

Theorem 1. *Sei T eine präkompakte Teilmenge eines Hilbertraumes. Dann gelten die folgenden Abschätzungen:*

(i). $N(T,\varepsilon) \preceq |\log\varepsilon|^{\beta}$ *impliziert* $\log N(\mathrm{aco}(T),\varepsilon) \preceq |\log\varepsilon|^{\beta+1}$ *für* $\beta > 0$.

(ii). $N(T,\varepsilon) \preceq \varepsilon^{-\alpha}|\log\varepsilon|^{-\beta}$ *impliziert* $\log N(\mathrm{aco}(T),\varepsilon) \preceq \varepsilon^{-\frac{2\alpha}{2+\alpha}}|\log\varepsilon|^{-\frac{2\beta}{2+\alpha}}$ *für* $\alpha > 0$ *und* $\beta \in \mathbb{R}$.

(iii). $\log N(T,\varepsilon) \preceq \varepsilon^{-\alpha}|\log\varepsilon|^{-\beta}$ *impliziert*
$\log N(\mathrm{aco}(T),\varepsilon) \preceq \varepsilon^{-2}|\log\varepsilon|^{1-2/\alpha}(\log|\log\varepsilon|)^{-2\beta/\alpha}$
für $0 < \alpha < 2$ *und* $\beta \in \mathbb{R}$.

(iv). $\log N(T,\varepsilon) \preceq \varepsilon^{-2}|\log\varepsilon|^{-\beta}$ *impliziert* $\log N(\mathrm{aco}(T),\varepsilon) \preceq \varepsilon^{-2}(\log|\log\varepsilon|)^{2-\beta}$ *für* $\beta > 2$.

(v). $\log N(T,\varepsilon) \preceq \varepsilon^{-2}|\log\varepsilon|^{-2}$ *impliziert* $\log N(\mathrm{aco}(T),\varepsilon) \preceq \varepsilon^{-2}(\log|\log\varepsilon|)^{2}$.

(vi). $\log N(T,\varepsilon) \preceq \varepsilon^{-2}|\log\varepsilon|^{-\beta}$ *impliziert* $\log N(\mathrm{aco}(T),\varepsilon) \preceq \varepsilon^{-2}|\log\varepsilon|^{2-\beta}$ *für* $\beta < 2$.

(vii). $\log N(T,\varepsilon) \preceq \varepsilon^{-\alpha}$ *impliziert* $\log N(\mathrm{aco}(T),\varepsilon) \preceq \varepsilon^{-\alpha}$ *für* $\alpha > 2$.

Beachten wir, dass (ii) eine Verbesserung von Dudley (1987) durch Carl (1997) ist, wobei dieses Resultat tatsächlich bereits in Theorem 5.10.1 in Carl and Stephani (1990) enthalten ist. Abschätzungen (i),(iii) und (vii) sind in Carl, Kyrezi, and Pajor (1999) enthalten. Die *kritischen Fälle* $\log N(T,\varepsilon) \preceq \varepsilon^{-2}|\log\varepsilon|^{-\beta}$, $\beta \in \mathbb{R}$, sind vergleichsweise schwierig zu handhaben. Ein erstes Resultat betrifft den Fall $\beta = 0$ und stammt von Gao (2001), der ebenso ein geistreiches Beispiel für eine Menge ersann, die zeigte, dass seine obere Schranke bestmöglich ist. Gaos Ideen wurden von Creutzig and Steinwart (2002) aufgegriffen und auf den Fall $\beta < 2$ sowie auf Banachräume vom Typ $p \in (1,2]$ ausgedehnt. Seither war es eine offene Frage, was in den Fällen $\beta \geq 2$, also Teilen (iv) und (v) von Theorem 1, geschieht. Wir werden zeigen, dass es uns kürzliche Entwicklungen in der Theorie der kleinen Abweichungen durch Aurzada and Lifshits (2008) ermöglichen, eine obere Schranke im Fall $\beta > 2$ zu finden: Wir konstruieren in Abhängigkeit von T eine Gaußsche Zufallsvariable X mit Werten in einem separablen Banachraum $E \subseteq \mathbb{R}^T$. Über das Resultat aus Aurzada and Lifshits (2008), eine Ausdehnung der klasischen "Talagrand lower bound", erhalten wir aus einer oberen Schranke für $\log N(T,\varepsilon)$ eine obere Schranke für die Funktion der kleinen Abweichungen

$$\phi(\varepsilon) := -\log \mathbb{P}\left[\|X\|_E \leq \varepsilon\right].$$

Diese mündet durch Anwendung einer der bekannten Kuelbs-Li Ungleichungen in eine obere Schranke für die metrische Entropie $\log N(K,\varepsilon)$ der Einheitskugel K des reproduzierenden Kernhilbertraumes von X. Unser Vorgehen wird von einer engen Verbindung zwischen der metrischen Entropie von K und $\mathrm{aco}(T)$ ergänzt, die in Abschnitt 3.5 fundiert wird, vgl. auch Gao (2004). Es stellt sich heraus, dass der oben beschriebene probabilistische Zugang für alle Fälle von Theorem 1 funktioniert, in denen Dudleys Integral

$$\int_0^\infty \sqrt{\log N(T,\varepsilon)}d\varepsilon \tag{1.1}$$

konvergiert, also für Fälle (i) bis (iv).
Die verbleibenden Abschätzungen (v) bis (vii) können aus Remark 5.3 in Carl, Kyrezi, and Pajor (1999) abgeleitet werden, wie Lifshits (2009) beobachtet hat, der zudem einen Beweis gab, welcher sich auf Gaußprozesse und die Sudakov-Minoration gründet, siehe Sudakov (1969). Bemerkenswerterweise helfen probabilistische Methoden auch in Fällen, in denen das Dudley Integral nicht mehr notwendig konvergiert.
Insgesamt erhalten wir einen kurzen, einfachen und probabilistischen Beweis von Theorem 1 mit einheitlicher Behandlung der Fälle (i) bis (iv) sowie (v) bis (vii). Dieser ist in Abschnitt 4.1 ausgeführt.
Außerdem geben wir einen direkten Beweis für Fall (v) in Abschnitt 4.2, der sich auf die ursprüngliche Argumentation in Gao (2001) sowie Creutzig and Steinwart (2002) gründet.
Zusätzlich werden wir die Teile (iv) bis (vii) beweisen, indem wir Methoden verwenden, die ihren Ursprung in der Theorie der majorisierenden Maße haben, siehe Abschnitt 4.3.
Wir möchten unterstreichen, dass unabhängig die Abschätzungen (iv) und (v) von Carl, Hinrichs, and Pajor (2010) mit anderen Methoden gefunden wurden. Es ist jedoch unbekannt, ob (iv) und (v) bestmöglich sind. Dies bleibt ein offenes Problem. Hingegen kann der existierenden Literatur entnommen werden, dass Abschätzungen (i), (ii), (iii), (vi) und (vii) bestmöglich sind.
Kapitel 5 enthält Resultate, die im Bereich der kritisch großen Entropie liegen, in gleicher Weise wie in Kapitel 4 werden wir lebhaft von der Verbindung zwischen der Einheitskugel des reproduzierenden Kernhilbertraums einer Gaußschen Zufallsvariable und $\mathrm{aco}(T)$ Gebrauch machen.
In Abschnitt 5.1 stellen wir unter Zuhilfenahme einer Kuelbs-Li Ungleichung untere Schranken für die kleinen Abweichungen bereit, wenn eine untere Schranke für die Entropie der Einheitskugel des zugehörigen reproduzierenden Kernhilbertraumes gegeben ist. Diese Abschätzungen werden im folgenden dazu verwendet, die Konvergenzraten der Funktion der kleinen Abweichungen von Folgen unabhängiger Gaußscher Zufallsvariablen sowie eines Volterra-Typ Prozesses zu identifizieren.
In Abschnitt 5.2 beschäftigen wir uns mit Optimalitätsbetrachtungen bezüglich der Kuelbs-Li Ungleichungen. Es folgt aus Kuelbs and Li (1993) für eine Funktion f mit $f(\varepsilon) \preceq \phi(\varepsilon)$ die Beziehung

$$\log N(K, \varepsilon/\sqrt{f(\varepsilon)}) \preceq f(\varepsilon), \tag{1.2}$$

während ebenso

$$\log N(K, \varepsilon/\sqrt{2\phi(\varepsilon)}) \succeq \phi(2\varepsilon) \tag{1.3}$$

Gültigkeit besitzt. Die erste Verbindung (1.2) verwenden wir, um eine obere Schranke für die metrische Entropie $\log N(K, \varepsilon)$ zu bestimmen, wenn eine obere Schranke für die Funktion der kleinen Abweichungen $\phi(\varepsilon)$ gegeben ist.

Ebenso erstrebenswert wäre es, aus einer unteren Schranke für $\phi(\varepsilon)$ eine untere Schranke für $\log N(K, \varepsilon)$ herleiten zu können. In Abschnitt (1.2) untersuchen wir, inwieweit dies möglich ist unter Verwendung von Beziehung (1.3). Dabei beschränkt insbesondere die Zahl zwei auf der rechten Seite von (1.3) den Anwendungsbereich auf Probleme, bei denen $\phi(\varepsilon)$ die Dopplungsbedingung $\phi(2\varepsilon) \succeq \phi(\varepsilon)$ erfüllt, was in vielen Fällen nicht gilt. Es würde auch keine Verbesserung bedeuten, $\phi(2\varepsilon)$ in (1.3) durch $\phi((1+\delta)\varepsilon)$ für ein festes $\delta > 0$ zu ersetzen; schließlich ist $\phi(2\varepsilon) \succeq \phi(\varepsilon)$ erfüllt, sobald $\phi((1+\delta)\varepsilon) \succeq \phi(\varepsilon)$ erfüllt ist. Kurzum, es wäre wünschenswert, die Relation

$$\log N(K, \varepsilon/\sqrt{2\phi(\varepsilon)}) \succeq \phi(\varepsilon)$$

zu beweisen. Wir zeigen jedoch, dass dies nicht möglich ist.

Die zweite Frage betrifft das Hauptresultat von Li and Linde (1999), welches darin besteht, die Ergebnisse von Kuelbs and Li (1993) so auszudehnen, dass eine eineindeutige Korrespondenz zwischen der Funktion der kleinen Abweichungen und der metrischen Entropie der Einheitskugel des zugehörigen reproduzierenden Kernhilbertraums bewiesen wurde, die sich in der Äquivalenz

$$\phi(\varepsilon) \approx \varepsilon^{-\alpha} |\log \varepsilon|^{\beta} \ \text{ gdw } \ \log N(K, \varepsilon) \approx \varepsilon^{-2\alpha/(2+\alpha)} |\log \varepsilon|^{2\beta/(2+\alpha)}, \tag{1.4}$$

mit $\alpha > 0$ und $\beta \in \mathbb{R}$ ausdrückt. In Abschnitt 5.2 werden wir zeigen, dass eine solch enge Beziehung im kritischen Fall nicht mehr gilt, d.h. wenn $\phi(\varepsilon)$ super-reguläres Verhalten zeigt oder $\log N(K, \varepsilon) \approx \varepsilon^{-2} J(1/\varepsilon)$ mit einer fallenden Funktion J, die im Unendlichen langsam variiert.

Schließlich diskutieren wir die Gao-Menge aus Gao (2001), die leichte Modifikationen erfährt. Mit Hilfe von Operatorzahlen, welche der Theorie des Nicht-Determinismus stochastischer Prozesse entstammen, siehe Linde (2008), geben wir untere Schranken für die Entropie der absolutkonvexen Hüllen jener speziellen Mengen, die in einigen Fällen besser sind als die gemeinhin bekannten, jedoch trotzdem nicht ausreichen, um zu zeigen, dass die Abschätzungen (iv) bzw. (v) bestmöglich sind. Dies bleibt ein offenes Problem.

2. INTRODUCTION

The leitmotif of this work can be described in quite a simple manner: Given three distinct points in the plane, one needs at most three disks of arbitrary radius to cover them. Those three points may be regarded to constitute the vertices of a triangle. Obviously, many more than three disks are necessary to cover the whole figure, and their number increases as their radii decrease. The triangle is the convex hull of its vertices. In this sense the question can be generalized: How many balls of certain radius are needed to cover the convex hull of a set of points in some linear normed space provided information about how many balls are needed to cover the original set? This issue in all its facets pervades the whole work. We formulate the setting precisely:

Let (E,d) be any metric space, $T \subseteq E$ be a precompact subset and $B(x_0,\varepsilon) := \{x \in E : d(x_0,x) \le \varepsilon\}$. The **covering numbers** of T are defined by

$$N(T,\varepsilon) := \min\{n \in \mathbb{N} : \exists\, t_1,...,t_n \in E \text{ such that } T \subseteq \bigcup_{i=1}^{n} B(t_i,\varepsilon)\}.$$

We will refer to $\log N(T,\varepsilon)$ as **metric entropy** of T. In the following, T will denote a subset of a Hilbert space and will be equipped with the metric induced by the inner product. We are only concerned with *separable* Hilbert spaces, even if not explicitly stated.

The main task is to obtain bounds for the covering numbers of the convex hull of T relying on information about the set T.

In both Carl (1982) and Ball and Pajor (1991) the entropy of the convex hull of sets with "few" vectors is studied, whereas their norms or at least bounds for these are known. Dudley (1987) was the first to generalize the question in such a way that only bounds for covering numbers of the original set T were given, without special knowledge about the norms of its (extreme) points.
He proves that $N(T,\varepsilon) \le c_1\varepsilon^{-\alpha}$ implies $\log N(\mathrm{co}(T),\varepsilon) \le c_2\varepsilon^{-\gamma}$ for $\alpha > 0$

and all $\gamma > \frac{2\alpha}{2+\alpha}$, where c_1, c_2 are constants independent of ε.
His research was motivated by applications in the field of empirical processes, and furthermore, entropy of convex hulls found applications – see Carl, Kyrezi, and Pajor (1999)– in various branches of mathematics such as analysis, approximation theory, geometry as well as probability, cf. Carl and Stephani (1990), Ledoux and Talagrand (1991), Edmunds and Triebel (1996), Pisier (1989). Nevertheless, let us underline how natural and intuitive the task is, which makes considerations interesting disregarding application possibilities.
Dudley's assertion cited above is nowadays accompanied by a number of statements for different rates of growth of $N(T,\varepsilon)$ as well as there were several extensions to the Banach space case, for the latter cf. Carl, Kyrezi, and Pajor (1999), Steinwart (2000), Creutzig and Steinwart (2002) and Steinwart (2004).
Instead of working with the convex hull, we will prefer the **absolutely convex hull** of T

$$\mathrm{aco}(T) := \{\sum_{i=1}^{n} \lambda_i t_i : \ \sum_{i=1}^{n} |\lambda_i| = 1, \ t_1, ..., t_n \in T, \ n \in \mathbb{N}\}$$

which enables us to deal with duality.
For functions f, g we use the notation $f \preceq g$ if $\limsup_{x\to a} \frac{f(x)}{g(x)} < +\infty$ with $a \in [0, +\infty]$. In this case, we also write $g \succeq f$. Moreover, we use $f \approx g$ if $f \preceq g$ and $g \preceq f$. For sequences we employ the analogous notation.
Let us concentrate the statements for the most important rates of $N(T,\varepsilon)$ in the following theorem.

Theorem 2. *Let T be a precompact subset of some Hilbert space. Then the following upper estimates are valid:*

(i). $N(T,\varepsilon) \preceq |\log\varepsilon|^{\beta}$ *implies* $\log N(\mathrm{aco}(T),\varepsilon) \preceq |\log\varepsilon|^{\beta+1}$ *for* $\beta > 0$.

(ii). $N(T,\varepsilon) \preceq \varepsilon^{-\alpha}|\log\varepsilon|^{-\beta}$ *implies* $\log N(\mathrm{aco}(T),\varepsilon) \preceq \varepsilon^{-\frac{2\alpha}{2+\alpha}}|\log\varepsilon|^{-\frac{2\beta}{2+\alpha}}$ *for* $\alpha > 0$ *and* $\beta \in \mathbb{R}$.

(iii). $\log N(T,\varepsilon) \preceq \varepsilon^{-\alpha}|\log\varepsilon|^{-\beta}$ *implies*
$\log N(\mathrm{aco}(T),\varepsilon) \preceq \varepsilon^{-2}|\log\varepsilon|^{1-2/\alpha}(\log|\log\varepsilon|)^{-2\beta/\alpha}$
for $0 < \alpha < 2$ *and* $\beta \in \mathbb{R}$.

(iv). $\log N(T,\varepsilon) \preceq \varepsilon^{-2}|\log\varepsilon|^{-\beta}$ *implies* $\log N(\mathrm{aco}(T),\varepsilon) \preceq \varepsilon^{-2}(\log|\log\varepsilon|)^{2-\beta}$ *for* $\beta > 2$.

(v). $\log N(T,\varepsilon) \preceq \varepsilon^{-2}|\log\varepsilon|^{-2}$ *implies* $\log N(\mathrm{aco}(T),\varepsilon) \preceq \varepsilon^{-2}(\log|\log\varepsilon|)^2$.

(vi). $\log N(T,\varepsilon) \preceq \varepsilon^{-2}|\log\varepsilon|^{-\beta}$ *implies* $\log N(\mathrm{aco}(T),\varepsilon) \preceq \varepsilon^{-2}|\log\varepsilon|^{2-\beta}$ *for* $\beta < 2$.

(vii). $\log N(T,\varepsilon) \preceq \varepsilon^{-\alpha}$ *implies* $\log N(\mathrm{aco}(T),\varepsilon) \preceq \varepsilon^{-\alpha}$ *for* $\alpha > 2$.

Note that (ii) is an improvement of Dudley (1987) by Carl (1997), where in fact the result is already included in Theorem 5.10.1 in Carl and Stephani (1990). Estimates (i), (iii) and (vii) are contained in Carl, Kyrezi, and Pajor (1999). The *critical cases* $\log N(T,\varepsilon) \preceq \varepsilon^{-2}|\log\varepsilon|^{-\beta}$ for $\beta \in \mathbb{R}$ have been comparatively difficult to treat. The first result was for $\beta = 0$ and due to Gao (2001), who also invented an ingenious example to prove that the estimate is best possible. Creutzig and Steinwart (2002) extended his ideas to the case $\beta < 2$ and to Banach spaces of type $p \in (1,2]$.
It has been an open question for almost ten years what happens in the remaining cases $\beta \geq 2$, i.e. cases (iv) and (v). We will show that very recent developments in the theory of small deviations, see Aurzada and Lifshits (2008), enable us to find a reasonable upper bound for the case of $\beta > 2$. For this, we construct a Gaussian random variable X with values in a separable Banach space $E \subseteq \mathbb{R}^T$. With the result from Aurzada and Lifshits (2008), an extension of the classical "Talagrand lower bound", we derive an upper bound for the **small deviations function**

$$\phi(\varepsilon) := -\log \mathbb{P}\left[\|X\|_E \leq \varepsilon\right] \tag{2.1}$$

from an upper bound for $\log N(T,\varepsilon)$. This upper bound for $\phi(\varepsilon)$ runs into an upper bound for the metric entropy of the unit ball of the reproducing kernel Hilbert space of X with the aid of one of the famous Kuelbs-Li inequalities. This procedure will be finally coupled with a precise link between the reproducing kernel Hilbert space (short: rkHs) of X and $\mathrm{aco}(T)$ in terms of entropy, see the preliminary chapter, section 3.5, cf. also Gao (2004). It turns out that our probabilistic approach works for all cases of Theorem 2 where Dudley's integral $\int_0^\infty \sqrt{\log N(T,\varepsilon)}d\varepsilon$ converges, i.e. cases (i) to (iv). The remaining cases (v) to (vii) may be derived from Remark 5.3 in Carl, Kyrezi, and Pajor (1999). This obseration is due to Lifshits (2009) who also gave a proof relying on Gaussian processes and Sudakov minoration, see Sudakov (1969). Notably, this shows that the probabilistic method has not reached the end of the line, even in cases of not necessarily converging Dudley integral. Therefore, we get quite a short, simplified, and probabilistic proof of all cases of Theorem 2 with unified treatment of (i) to (iv) and (v)

to (vii), respectively.
As a supplement, we give a direct proof of (v) based on the original arguments of both Gao (2001) and Creutzig and Steinwart (2002). In addition, we give another proof of Theorem 2, parts (iv)–(vii) by employing methods originating in the theory of majorizing measures which were mainly developed in Li and Linde (2000) based on Bühler (1999) and Bühler, Li, and Linde (2001). Cases (iv) and (v) have not appeared in the literature yet. However, let us point out that both have been found independently by Carl, Hinrichs, and Pajor (2010) using other techniques than we will. They compute the same upper bounds as we do, but no one knows whether the estimates (iv) and (v) are best possible. This remains an open problem. Let us remark that estimates (i), (ii), (iii), (vi), (vii) are best possible.
Chapter 5 contains results in the range of critically large entropy. In the same way as in chapter 4, we will vividly use the link between the rkHs of Gaussian processes and absolutely convex hulls. In section 5.1 we provide lower bounds for the small deviations given lower bounds for entropy of the unit ball of rkHs as a consequence of Kuelbs-Li inequality. These bounds are then applied to identify the correct orders of small deviations of Gaussian independent sequences as well as those of a Volterra type process.
Section 5.2 is devoted to optimality considerations concerning the Kuelbs-Li inequalities. It follows from Kuelbs and Li (1993) that on the one hand for each function f with $\phi(\varepsilon) \preceq f(\varepsilon)$ the relation

$$\log N(K, \varepsilon/\sqrt{f(\varepsilon)}) \preceq f(\varepsilon). \tag{2.2}$$

is valid, while we have

$$\log N(K, \varepsilon/\sqrt{2\phi(\varepsilon)}) \succeq \phi(2\varepsilon) \tag{2.3}$$

on the other hand. The first connection (4.1) is suitable for computing an upper bound for the metric entropy $\log N(K, \varepsilon)$ of the unit ball K of the corresponding rkHs, if an upper bound for the small ball function $\phi(\varepsilon)$ is given.
Clearly, it would be convenient to have an inverse result, i.e. we want to establish a lower bound for $\log N(K, \varepsilon)$ given a lower bound of $\phi(\varepsilon)$. In this section we investigate the question, to what extend this is possible. The key step towards this is the relation (2.3). Unfortunately, the number two on the right hand side of (2.3) significantly restricts the scope of application to problems, where $\phi(\varepsilon)$ fulfills the doubling condition $\phi(2\varepsilon) \succeq \phi(\varepsilon)$. Note that in many cases $\phi(\varepsilon)$ does not satisfy that condition. We observe that it would not be an improvement to replace $\phi(2\varepsilon)$ by $\phi((1+\delta)\varepsilon)$ for a fixed

$\delta > 0$ since $\phi(2\varepsilon) \succeq \phi(\varepsilon)$ is fulfilled whenever $\phi((1+\delta)\varepsilon) \succeq \phi(\varepsilon)$ is. In other words, it would be desirable to prove

$$\log N(K, \varepsilon/\sqrt{2\phi(\varepsilon)}) \succeq \phi(\varepsilon). \tag{2.4}$$

However, we will show that this is not possible.
The second question affects a major result of the article Li and Linde (1999) which is to supplement the findings of Kuelbs and Li (1993) by establishing a one-to-one correspondence between the small deviations and the metric entropy of the unit ball of the rkHs of a Gaussian random variable which may be expressed through the equivalence

$$\phi(\varepsilon) \approx \varepsilon^{-\alpha} |\log \varepsilon|^{\beta} \ \text{ iff } \ \log N(K, \varepsilon) \approx \varepsilon^{-2\alpha/(2+\alpha)} |\log \varepsilon|^{2\beta/(2+\alpha)}, \tag{2.5}$$

where $\alpha > 0$ and $\beta \in \mathbb{R}$.
In the second part of section 5.2 , we will show that such a tight relation does not hold anymore in the critical case, i.e., if $\phi(\varepsilon)$ shows super-regular behaviour or $\log N(K, \varepsilon) \approx \varepsilon^{-2} J(1/\varepsilon)$ where J denotes a decreasing function slowly varying at infinity.
We finish with a discussion on Gao set, which will be slightly modified and generalized. With the help of operator numbers arising from probability, especially from the theory of non-determinism of stochastic processes, we give lower bounds for the entropy of the absolutely convex hull of these particular sets which are in some cases better than the known ones, though apart from showing the sharpness of Theorem 2 in cases (iv) or (v). This remains an open question.

3. PRELIMINARIES

In this chapter, we introduce the main quantities of this work, cite important results and useful assertions already having appeared in the literature or are known, at least to experts. We restrict ourselves to give a tailored overview fitting to our considerations, which mainly take place in the setting of separable Hilbert and Banach spaces, in particular in section 3.2, though there is a huge general theory behind, cf. Bogachev (1998). Let us point out that we do not claim originality of the results, even if we give proofs instead of exact references, which are sometimes hardly to be found.

3.1 Entropy and covering numbers

The concept of covering numbers $N(T, \varepsilon)$ was introduced in Kolmogorov and Tikhomirov (1959). In particular, $N(T, \varepsilon)$ may be regarded as a function in ε mapping $(0, \infty)$ into $\mathbb{N}$ if T is precompact. If $N(T, \varepsilon)$ increases exponentially as ε tends to 0, it is common to regard the quantity

$$\log N(T, \varepsilon)$$

which is often referred to as **metric entropy** of T.

The following is a concept inverse to covering numbers. The **n−th entropy number** $\varepsilon_n(T)$ is defined by

$$\varepsilon_n(T) := \inf\{\varepsilon > 0 |\ N(T, \varepsilon) \leq n\}.$$

and the **n−th dyadic entropy number** $e_n(T)$ by

$$e_n(T) := \varepsilon_{2^{n-1}}(T).$$

The entropy numbers $e_n(u)$ of a linear operator $u : E \to F$ between two Banach spaces are defined by

$$e_n(u) := e_n(u(B_E)),$$

where B_E denotes the closed unit ball of E; see Carl and Stephani (1990) for various properties and further information about entropy numbers. We will use both quantities, covering numbers as well as entropy numbers. Hence it will be convenient to be able to switch from one to the other. Unfortunately, we could not find any literature to cite concerning their relations. We begin with stating two easy but clarifying lemmata.

Lemma 3. *Let $[T, d]$ be a precompact metric space, $\varepsilon > 0$ and $n \in \mathbb{N}$.*

(i). $\varepsilon_n(T) < \varepsilon$ *implies* $N(T, \varepsilon) \leq n$.

(ii). $\varepsilon_n(T) > \varepsilon$ *implies* $N(T, \varepsilon) \geq n + 1$.

(iii). $N(T, \varepsilon) \leq n$ *implies* $\varepsilon_n(T) \leq \varepsilon$.

(iv). $N(T, \varepsilon) \geq n + 1$ *implies* $\varepsilon_n(T) \geq \varepsilon$.

Lemma 4. *Let $[T, d]$ be a precompact metric space, $\varepsilon > 0$ and $n \in \mathbb{N}$.*

(i). $e_{n+1}(T) < \varepsilon$ *implies* $\log N(T, \varepsilon) \leq n \log 2$.

(ii). $e_{n+1}(T) > \varepsilon$ *implies* $\log N(T, \varepsilon) \geq n \log 2$.

(iii). $\log N(T, \varepsilon) \leq n \log 2$ *implies* $e_{n+1}(T) \leq \varepsilon$.

(iv). $\log N(T, \varepsilon) \geq n \log 2$ *implies* $e_n(T) \geq \varepsilon$.

Next, we formulate connections between entropy numbers and covering numbers, which will be used freely throughout this work.

Proposition 5. *Let $[T, d]$ be a precompact metric space, $\alpha > 0$ and $\beta \in \mathbb{R}$.*

(i). $\varepsilon_n(T) \preceq n^{-\alpha}(\log(n+1))^{\beta}$ *iff* $N(T, \varepsilon) \preceq \varepsilon^{-\frac{1}{\alpha}} |\log \varepsilon|^{\frac{\beta}{\alpha}}$.

(ii). $\varepsilon_n(T) \succeq n^{-\alpha}(\log(n+1))^{\beta}$ *iff* $N(T, \varepsilon) \succeq \varepsilon^{-\frac{1}{\alpha}} |\log \varepsilon|^{\frac{\beta}{\alpha}}$.

(iii). $e_n(T) \preceq n^{-\alpha}(\log(n+1))^{\beta}$ *iff* $\log N(T, \varepsilon) \preceq \varepsilon^{-\frac{1}{\alpha}} |\log \varepsilon|^{\frac{\beta}{\alpha}}$.

(iv). $e_n(T) \succeq n^{-\alpha}(\log(n+1))^{\beta}$ *iff* $\log N(T, \varepsilon) \succeq \varepsilon^{-\frac{1}{\alpha}} |\log \varepsilon|^{\frac{\beta}{\alpha}}$.

(v). $e_n(T) \preceq n^{-1/2}(\log(n+1))^{1/2-\alpha}(\log\log(n+2))^{\beta}$ *iff* $\log N(T, \varepsilon) \preceq \varepsilon^{-2} |\log \varepsilon|^{1-2\alpha}(\log|\log \varepsilon|)^{2\beta}$.

(vi). $e_n(T) \succeq n^{-1/2}(\log(n+1))^{1/2-\alpha}(\log\log(n+2))^{\beta}$ *iff* $\log N(T, \varepsilon) \succeq \varepsilon^{-2} |\log \varepsilon|^{1-2\alpha}(\log|\log \varepsilon|)^{2\beta}$.

Proof. We restrict ourselves to the proof of (i) since the other assertions are handled analogously.

Let $F(x) := x^{-\frac{1}{\alpha}}|\log x|^{\frac{\beta}{\alpha}}$ which is monotone for $0 < x < c_F := \min\{1, \exp[\beta]\}$ and $G(y) := y^{-\alpha}\log(y+1)^{\beta}$ which is monotone for $y > c_G := \max\{1, \exp[\frac{\beta}{\alpha}]\}$. Moreover, there is an $n_F \in \mathbb{N}$ so that we have $n^{-\alpha}(\log(n+1))^{\beta} < c_F$ for all $n \geq n_F$.

From left to right: We know that there is a constant $c > 0$ so that $\varepsilon_n(T) \leq cn^{-\alpha}(\log(n+1))^{\beta}$ for all $n \in \mathbb{N}$. For $\varepsilon > 0$ with

$$0 < \varepsilon < \min\{(c_G+1)^{-\alpha}(\log(c_{G+2}))^{\beta}, (n_F+1)^{-\alpha}(\log(n_F+2))^{\beta}\}$$

there is an $n(\varepsilon) \geq n_F + 1$ so that

$$\varepsilon_{n(\varepsilon)+1}(T) \leq c(n(\varepsilon)+1)^{-\alpha}(\log(n(\varepsilon)+2))^{\beta} \leq c\varepsilon < cn(\varepsilon)^{-\alpha}(\log n(\varepsilon)+1)^{\beta}. \tag{3.1}$$

Applying the function F to (3.1) yields

$$F(n(\varepsilon)^{-\alpha}(\log n(\varepsilon))^{\beta}) \leq F(\varepsilon)$$

and therefore

$$n(\varepsilon)+1 \leq \frac{n(\varepsilon)+1}{F(n(\varepsilon)^{-\alpha}(\log n(\varepsilon))^{\beta})}F(\varepsilon). \tag{3.2}$$

In view of Lemma 3(i), (3.1) and (3.2) we get

$$N(T, c\varepsilon) \leq c\frac{n(\varepsilon)+1}{F(n(\varepsilon)^{-\alpha}(\log n(\varepsilon))^{\beta})}\varepsilon^{-\frac{1}{\alpha}}|\log\varepsilon|^{\frac{\beta}{\alpha}}.$$

Since by (3.1), $n(\varepsilon) \to \infty$ as $\varepsilon \to 0$ and

$$\limsup_{n\to\infty} \frac{n}{F(n^{-\alpha}(\log n)^{\beta})} < +\infty$$

as one computes, the assertion follows.

From right to left: According to the assumption, there is a constant $c > 0$ so that for all $\varepsilon > 0$ we have

$$N(T, \varepsilon) < c\varepsilon^{-\frac{1}{\alpha}}|\log\varepsilon|^{\frac{\beta}{\alpha}}.$$

Let $n > c_G + 1$. There is an $\varepsilon(n) > 0$ so that

$$n - 1 \leq \varepsilon(n)^{-\frac{1}{\alpha}} |\log \varepsilon(n)|^{\frac{\beta}{\alpha}} \leq n. \tag{3.3}$$

and

$$N(T, \varepsilon(n)) < c\varepsilon(n)^{-\frac{1}{\alpha}} |\log \varepsilon(n)|^{\frac{\beta}{\alpha}} \leq cn \tag{3.4}$$

We apply G to (3.3) which yields

$$G(\varepsilon(n)^{-\frac{1}{\alpha}} |\log \varepsilon(n)|^{\frac{\beta}{\alpha}}) \leq G(n-1).$$

With Lemma 3(iii)in view of (3.4) we now achieve

$$\varepsilon_{\lceil cn \rceil}(T) \leq \varepsilon(n) \leq c' \frac{\varepsilon(n)}{G(\varepsilon(n)^{-\frac{1}{\alpha}} |\log \varepsilon(n)|^{\frac{\beta}{\alpha}})} n^{-\alpha} (\log(n+1))^{\beta}.$$

By (3.3), $\varepsilon(n) \to 0$ as $n \to \infty$. Since one computes

$$\limsup_{\varepsilon \to 0} \frac{\varepsilon}{G(\varepsilon^{-\frac{1}{\alpha}} |\log \varepsilon|^{\frac{\beta}{\alpha}})} < +\infty,$$

the proof of (i) is completed. □

3.2 Gaussian processes, random variables, and measures

Let T be a non-empty set and $(\Omega, \mathcal{F}, \mathbb{P})$ a complete probability space. As usual, $\mathfrak{B}(E)$ denotes the Borel-σ-algebra of some topological space E. A family

$$X = (X_t)_{t \in T}$$

of real-valued random variables

$$X_t : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathfrak{B}(\mathbb{R}))$$

is called a **stochastic process.**

Such a family $X = (X_t)_{t \in T}$ will be called **Gaussian stochastic process** or **Gaussian process** for short, if the linear combination

$$\sum_{i=1}^{n} \lambda_i X_{t_i} \tag{3.5}$$

is normally distibuted for all choices of $\lambda_1, \cdots, \lambda_n \in \mathbb{R}$, $t_1, \cdots, t_n \in T$ and $n \in \mathbb{N}$.

For each fixed $\omega \in \Omega$, $X_{\cdot}(\omega)$ is a real valued function on T called **path** of the process. It is common to think of the stochastic process X as a mapping

$$X : \Omega \to \mathbb{R}^T.$$

We will mainly be concerned with the case, where X has paths a.s. in a separable Banach space $E \subseteq \mathbb{R}^T$. Henceforth, that will be the setting.

If the mapping $X : \Omega \to E$ is $(\mathcal{F}, \mathfrak{B}(E))-$measurable, we will call it **Borel random variable**. Such a Borel random variable will be a **Gaussian random variable**, Grv for short, if in addition

$$\langle X, a \rangle \tag{3.6}$$

is normally distributed for all a from the topological dual E^* of E. In this context, $\langle \cdot, \cdot \rangle$ denotes the duality between E and E^*.

Let us underline that throughout this work we only handle **centered** Gaussian processes and centered Gaussian random variables, which means that all the real valued random variables in (3.5) and (3.6) have expectation zero and hence are centered themselves. In respect thereof, centeredness will not be mentioned or demanded explicitly anymore.

As common, we will denote by $C(T)$ the space of all real-valued continuous functions defined on a compact metric space T and identify its topological dual with $M(T)$, the space of all regular Borel measures on T equipped with variation norm as justified by Riesz Representation Theorem. In particular, $M(T)$ contains all point measures δ_t for $t \in T$.
We denote by $\mathrm{extr}(A)$ the set of extreme points of a set A in some linear space. Following Roy (1987), a point $x \in A$ is called **extreme point**, if $x = \lambda y + (1 - \lambda)z$ for $0 < \lambda < 1$ and $y, z \in A$ implies $x = y = z$. For the following considerations we provide three classical theorems.

Theorem 6 (Arens-Kelly)**.** *Let T be a compact metric space, then* $\mathrm{extr}(B_{M(T)}) = \{\pm\delta_t : t \in T\}$.

Theorem 7 (Krein-Milman)**.** *Let C be a non-empty compact subset of a locally convex space. Then* $\mathrm{extr}(C)$ *is not empty. If C is also convex, then C is the closed convex hull of* $\mathrm{extr}(C)$.

Theorem 8 (Banach-Alaoglu). *If E is a normed linear space, then B_{E^*} is weak*-compact. If E is additionally separable, then the weak*–topology is metrizable on B_{E^*}.*

Combining all three theorems yields

$$B_{M(T)} = \overline{aco\{\delta_t :\ t \in T\}}^{wk*}. \tag{3.7}$$

We will make use of this important fact subsequently.

Proposition 9. *Let $X = (X_t)_{t \in T}$ be a Gaussian Process indexed by a compact metric space T and possessing paths in $C(T)$ a.s. Then X is a Gaussian random variable.*

Proof. We regard X to be a mapping from Ω into $C(T)$. Concerning measurability of X w.r.t. $\mathcal{F}$ and $\mathfrak{B}(E)$ we employ Theorem 2.1 of Vakhaniya, Tarieladze, and Chobanyan (1987) which states that

$$\sigma(\delta_t : t \in T) = \mathfrak{B}(E).$$

As conventional, $\sigma(\delta_t : t \in T)$ denotes the coarsest $\sigma-$algebra w.r.t. which all the point functionals δ_t are measurable. As a consequence, it suffices to know that X is measurable w.r.t. $\mathcal{F}$ and $\sigma(\delta_t : t \in T)$, but this is trivial.

Since X is assumed to be Gaussian, we know that all the linear combinations

$$\sum_{i=1}^{n} \lambda_i X_{t_i}$$

are normally distributed, so for each $\nu \in \operatorname{lin}\{\delta_t : t \in T\} \subseteq M(T)$ the real-valued random variable $\langle X, \nu \rangle$ is normally distributed.

Regarding (3.7) and enclosing the fact that the weak*–topology is metrizable on $B_{M(T)}$, we see that for $\mu \in M(T)$ there is a sequence $(\mu_n)_{n \in \mathbb{N}} \subseteq 2\|\mu\| \operatorname{aco}(\delta_t : t \in T)$ for which the relation

$$\lim_{n \to \infty} \langle f, \mu_n \rangle = \lim_{n \to \infty} \int_T f \ d\mu_n = \int_T f \ d\mu = \langle f, \mu \rangle.$$

is valid for each $f \in C(T)$. By Banach Steinhaus Theorem, the sequence $(\|\mu_n\|)_{n \in \mathbb{N}}$ is bounded, say by a constant $c > 0$. As a consequence and since X is in particular bounded and Gaussian, we can estimate

$$\sigma_n^2 := \mathbb{E}|\langle X, \mu_n \rangle|^2 \leq c\mathbb{E} \sup_{t \in T} |X(t)|^2 \leq c',$$

with a constant $c^{'} > 0$ independent of n, see Ledoux and Talagrand (1991). Hence, by Lebesgue's dominated convergence theorem we have

$$\lim_{n\to\infty} \mathbb{E}|\langle X, \mu_n\rangle|^2 = \mathbb{E}|\langle X, \mu\rangle|^2 =: \sigma^2.$$

The next step is to compute the characteristic functional

$$\begin{aligned}\widehat{\langle X, \mu\rangle}(u) &= \mathbb{E}\exp[iu\langle X, \mu\rangle] \\ &= \lim_{n\to\infty} \mathbb{E}\exp[iu\langle X, \mu_n\rangle] \\ &= \lim_{n\to\infty} \exp[-\frac{u^2}{2}\sigma_n^2] \\ &= \exp[-\frac{u^2}{2}\sigma^2],\end{aligned}$$

and it follows that $\langle X, \mu\rangle$ is normally distributed finishing the proof. □

For the sake of completeness and since used in the key paper Kuelbs and Li (1993), the notion of the third announced entity remains to be introduced. We name a probability measure γ on $(E, \mathfrak{B}(E))$ a **Gaussian measure** on the separable Banach space E if the image measures $\gamma \circ a^{-1}$ are normal distributions for all $a \in E^*$. Clearly, the image measure $\mathbb{P}\circ X^{-1}$ of a Gaussian random variable $X : \Omega \to E$ is a Gaussian measure. Bogachev (1998) gives an extensive overview on Gaussian measures.

3.3 Reproducing kernel Hilbert spaces

We want to introduce the notion of a reproducing kernel Hilbert space of a Gaussian random variable possessing values in a separable Banach space a.s. This approach is analogous to the construction of a rkHs of a Gaussian measure on a separable Banach space employed e.g. in Kuelbs and Li (1993) or Li and Linde (1999) based on Kuelbs (1976). In our exposition we also lean onto Dunker (1998) and Trenkmann (2009). For not to get confused with inner products, we write $a(X)$ instead of $\langle X, a\rangle$.

Let a Gaussian random variable X with values in a separable Banach space E a.s. be given. We define a Hilbert space H_X by

$$H_X := \overline{\{a(X) : \ a \in E^*\}}$$

as a subset of $L^2(\Omega, \mathbb{P})$ in its norm topology the closure is taken.

Further, let $S : H_X \to E$ be the operator defined via the Bochner integral by

$$Sg := \int_\Omega g(\omega)X(\omega)d\mathbb{P}(\omega).$$

Note that S is bounded by Fernique's Theorem on the integrability of a Grv, see Theorem 1.2.3 in Fernique (1997), as well as it is one-to-one by Proposition E.11 in Cohn (1980). The image $\mathcal{H}_X := S(H_X)$ of H_X under S equipped with the inner product

$$\langle Sg, Sh\rangle_{\mathcal{H}_X} := \langle g, h\rangle_{L^2}$$

is a Hilbert space which we call **reproducing kernel Hilbert space** of X. Its closed unit ball will be denoted by K_X or simply K if there is no doubt about the underlying Grv. It is always compact in the norm topolgy of E, see Kuelbs (1976). A useful characterization of $\mathcal{H}_X$ is formulated in the next Lemma.

Lemma 10. *For a Grv X with values in a separable Banach space E a.s. , the rkHs is represented by*

$$\mathcal{H}_X = \{x \in E \mid \exists c > 0 \forall\ a \in E^* : a(x)^2 \leq c^2 \int a(X)^2 d\mathbb{P}\}. \tag{3.8}$$

Moreover, $\|x\|_{\mathcal{H}_X} = \inf\{c > 0 |\ \forall a \in E^* :\ a(x)^2 \leq c^2 \int a(X)^2 d\mathbb{P}\}$.

Proof. Again with the help of Proposition E.11 in Cohn (1980) we see that

$$a(Sg)^2 = \left(\int ga(X)d\mathbb{P}\right)^2 \leq \int g^2 d\mathbb{P} \int a(X)^2 d\mathbb{P},$$

which immediately implies

$$\mathcal{H}_X \subseteq \{x \in E \mid \exists c > 0 \forall a \in E^* : a(x)^2 \leq c^2 \int a(X)^2 d\mathbb{P}\}$$

as well as $\inf\{c > 0 |\ \forall a \in E^* :\ a(x)^2 \leq c^2 \int a(X)^2 d\mathbb{P}\} \leq \|g\|_{L^2} = \|Sg\|_{\mathcal{H}_X}$.

For the inverse relations let $x \in E$ so that there is a constant $c > 0$ in such a way that for all $a \in E^*$ the inequality

$$a(x)^2 \leq c^2 \int a(X)^2 d\mathbb{P} \tag{3.9}$$

is valid. Then the functional $f_x(a(X)) := a(x)$ is well defined on $\{a(X) : a \in E^*\}$ which follows immediately from (3.9). It is additionally bounded, because

$$\|f_x\| = \sup_{\|a(X)\|=1} |a(x)| \leq c \left(\int a(X)^2 d\mathbb{P} \right)^{1/2} \leq c. \tag{3.10}$$

Hence, f_x can be continuously extended to H_X. By Riesz Theorem, there is an element $g \in H_X$ with $\|g\|_{L^2} = \|f_x\|$ so that

$$a(x) = \langle g, a(X) \rangle = \int g a(X) d\mathbb{P} = a \left(\int g X d\mathbb{P} \right).$$

It follows $x = \int gXd\mathbb{P}$ and (3.8) is proven. Since $\|x\|_{\mathcal{H}_X} = \|g\|_{L^2} = \|f_x\|$, also the second assertion follows in view of (3.10). □

We are now prepared to formulate the next Proposition which is crucial for our further considerations.

Proposition 11. *Let X be a Grv with values in a separable Banach space E and $u : H \to E$ a bounded linear operator on a Hilbert space H so that*

$$\int a(X)b(X)d\mathbb{P} = \langle u^*a, u^*b \rangle_H \tag{3.11}$$

for all $a, b \in E^*$. Then we have the equalities

$$u(H) = \mathcal{H}_X,\ u(B_H) = K_X \ \text{ and } \ \|g\|_{\mathcal{H}_X} = \inf\{\|h\|_H :\ u(h) = g\}.$$

Proof. Let $g \in E$ so that $g = u(h)$ for some $h \in H$. We can estimate

$$|a(g)|^2 = |\langle u^*a, h \rangle_H|^2 \leq \|h\|_H^2 \|u^*a\|_H^2 = \|h\|_H^2 \int a(X)^2 d\mathbb{P}$$

which gives $u(H) \subseteq \mathcal{H}_X$ as well as

$$\|g\|_{\mathcal{H}_X} = \inf\{c > 0 |\ \forall a \in E^* :\ a(g)^2 \leq c^2 \int a(X)^2 d\mathbb{P}\} \leq \|h\|_H. \tag{3.12}$$

For the reverse, let $g \in E$ fulfilling $a(g)^2 \leq c^2 \int a(X)^2 d\mathbb{P}$ for some $c > 0$ and all $a \in E^*$. We define a functional f_g on $H_0 = \{u^*a :\ a \in E^*\}$ by $f_g(u^*a) = a(g)$, and observe that f_g is well defined as for $u^*a = u^*b$

$$\begin{aligned}(a(g) - b(g))^2 &\leq c^2 \int ((a-b)(X))^2 d\mathbb{P} \\ &= c^2(\langle u^*a, u^*a \rangle - 2\langle u^*a, u^*b \rangle + \langle u^*b, u^*b \rangle) \\ &= 0\end{aligned}$$

by (3.11). Moreover,

$$|f_g(u^*a)|^2 = a(g)^2 \leq c^2 \int a(X)^2 d\mathbb{P} = c^2 \|u^*a\|_H^2$$

which shows that f_g is bounded, in particular $\|f_g\| = \|g\|_{\mathcal{H}_X}$. We then extend f_g continuously to $\overline{H_0}$ and define a functional $\tilde{f}_g$ on H by

$$\tilde{f}_g(h) := f_g(P_{\overline{H_0}} h)$$

where $P_{\overline{H_0}}$ denotes the orthogonal Projection from H to its closed subspace $\overline{H_0}$. Note that $\|f_g\| = \|\tilde{f}_g\|$. Of course, then there is an $h_g \in H$ with $\|\tilde{f}_g\| = \|h_g\|_H$ in such a way that $\tilde{f}_g(h) = \langle h, h_g \rangle_H$ for all $h \in H$. It follows

$$a(g) = f_g(u^*a) = \langle u^*a, h_g \rangle_H = a(uh_g)$$

for all $a \in E^*$ implying $g = uh_g$ which means $u(H) = \mathcal{H}_X$. Concerning the norm, we keep (3.12) in mind to conclude with $\|g\|_{\mathcal{H}_X} = \|f_g\| = \|h_g\|_H$ that

$$\|g\|_{\mathcal{H}_X} = \inf\{\|h\|_H : \ uh = g\}.$$

This equation will be of use for the last part of the assertion, namely $u(B_H) = K_X$. In a first step, we define an operator $\overline{u} : H_{/ker(u)} \to E$ by $\overline{u}(\overline{h}) := u(h)$, on the quotient Hilbert space $H_{/ker(u)}$ where h is a representant of $\overline{h}$. We observe that $\int a(X)b(X)d\mathbb{P} = \langle \overline{u}^*a, \overline{u}^*b \rangle$. Hence we have $\overline{u}(H_{/ker(u)}) = \mathcal{H}_X$. Since for each $g \in \mathcal{H}_X$ there is a unique $\overline{h}$ in $H_{/ker(u)}$ with $g = \overline{u}(\overline{h})$, we also have

$$\|g\|_{\mathcal{H}_X} = \|\overline{u}(\overline{h})\|_{\mathcal{H}_X} = \|\overline{h}\|_{H/ker(u)}.$$

As a consequence,

$$u(B_H) = \overline{u}(B_{H_{/ker(u)}}) = K_X,$$

which completes the proof. □

3.4 Small deviations

Let X be a Grv with values in a separable Banach space E a.s. The probability of small balls

$$\mathbb{P}\left[\|X\| \leq \varepsilon\right]$$

is regarded. We note that $\mathbb{P}\left[\|X\| \leq \varepsilon\right] > 0$ for all $\varepsilon > 0$. Suppose not: Then there would be an $\varepsilon_0 > 0$ with $\mathbb{P}\left[\|X\| \leq \varepsilon_0\right] = 0$. Because of separability, there is a countable set $D \subset E$ so that $E = \bigcup_{d \in D} B(d, \varepsilon_0)$. By Anderson inequality, see Theorem 2.8.10 in Bogachev (1998), $\mathbb{P}\left[\|X - d\| \leq \varepsilon_0\right] \leq \mathbb{P}\left[\|X\| \leq \varepsilon_0\right]$. This leads to

$$1 = \mathbb{P}\left[\|X\| < +\infty\right] \leq \sum_{d \in D} \mathbb{P}\left[\|X - d\| \leq \varepsilon_0\right] = 0$$

which is a contradiction.
A classical and well known statement refers to the small balls of Brownian motion $(B_t)_{t\in[0,1]}$, namely

$$\mathbb{P}\left[\sup_{t\in[0,1]} |B_t| < \varepsilon\right] \sim \frac{4}{\pi} \exp\left[-\frac{\pi^2}{8}\varepsilon^{-2}\right], \quad \varepsilon \to 0.$$

Since it is very difficult to give exact results, we usually restrict ourselves to finding the asymptotic rates on the log −level, i.e., one tries to evaluate

$$\phi(\varepsilon) := -\log \mathbb{P}\left[\|X\| \leq \varepsilon\right]$$

for $\varepsilon > 0$ tending to 0. The function ϕ will be called **small deviations function** of X. Although there is a clear dependence on the special choice of X, this shall only be reflected in notation of ϕ if there is any danger of confusion. We refer to the survey articles Li and Shao (2001) and Lifshits (1999) for historical development, various results and an overview on application areas such as empirical processes, convergence rates for functional laws of iterated logarithm or metric entropy of operators. The last will be of very interest in this work.
There are only a few general results for finding the rate of the small deviation function. Among them is a by now classical result, cf. Talagrand (1993), here cited in the form of Ledoux (1996).

Theorem 12 (Talagrand's lower bound)**.** *Let $(X_t)_{t\in T}$ be a Gaussian process and $\Psi(\varepsilon)$ be a function fulfilling $C_1\Psi(\varepsilon) \leq \Psi(\varepsilon/2), 0 < \varepsilon < \operatorname{diam}(T)$ and $\Psi(\varepsilon/2) \leq C_2\Psi(\varepsilon)$ with $C_2 > C_1 > 1$. Then $N(T, \varepsilon) \leq \Psi(\varepsilon)$ implies*

$$\log \mathbb{P}\left[\sup_{t,s\in T} |X_t - X_s| \leq \varepsilon\right] \geq -C\Psi(\varepsilon), \tag{3.13}$$

where $C > 0$ is a constant only depending on C_1 and C_2.

There are useful generalizations of Talagrand's lower bound in Aurzada and Lifshits (2008):

Theorem 13. *For $\beta > 0$, $N(T, \varepsilon) \preceq |\log \varepsilon|^\beta$ implies*

$$-\log \mathbb{P}\left[\sup_{t,s \in T} |X(t) - X(s)| \leq \varepsilon\right] \preceq |\log \varepsilon|^{\beta+1}.$$

Theorem 13 is a special case of Theorem 2 loc. cit., while the next one appears almost literally as Theorem 3.

Theorem 14. *Assume that $\log N(T, \varepsilon) \preceq \varepsilon^{-\alpha} |\log \varepsilon|^{-\beta}$ with $\beta > 2$ with some $0 < \alpha < 2, \beta \in \mathbb{R}$ or $\alpha = 2,\ \beta > 2$. Then*

$$\log \left|\log \mathbb{P}\left[\sup_{t,s \in T} |X_t - X_s| \leq \varepsilon\right]\right| \preceq \varepsilon^{-\frac{2\alpha}{2-\alpha}} |\log \varepsilon|^{-\frac{2\beta}{2-\alpha}},\ 0 < \alpha < 2 \tag{3.14}$$

and

$$\log\log \left|\log \mathbb{P}\left[\sup_{t,s \in T} |X_t - X_s| \leq \varepsilon\right]\right| \preceq \varepsilon^{-\frac{2}{\beta-2}},\ \alpha = 2 \tag{3.15}$$

Assuming $0 \in T$ and $X_0 = 0$ we have $\sup_{t \in T} |X_t| \leq \sup_{t,s \in T} |X_t - X_s|$. Hence, we can replace the supremum of increments by $\sup_{t \in T} |X_t|$ in (3.13) under these assumptions, as it will be done in the next section.

Another striking result is indebted to Kuelbs and Li (1993). They discovered a tight relation between the small ball function of a Grv X with values a separable Banach space E and the covering numbers of the unit ball K of its rkhs. This is expressed in the next Theorem.

Theorem 15. *Let X be a Grv with values in a separable Banach space a.s. and $\lambda > 0$. Then*

$$\log N(\lambda K, 2\varepsilon) \leq \frac{\lambda^2}{2} + \phi(\varepsilon), \tag{3.16}$$

$$\log N(\lambda K, \varepsilon) - \phi(2\varepsilon) \geq \log \Phi(\lambda + \alpha_\varepsilon), \tag{3.17}$$

whereas $\Phi(t) := (2\pi)^{-1/2} \int_{-\infty}^t \exp[-x^2/2]dx$ and $\Phi(\alpha_\varepsilon) \overset{!}{=} \mathbb{P}[\|X\| \leq \varepsilon]$.

We will return to Theorem 15 in section 4.1

3.5 Connections between reproducing kernel Hilbert spaces and absolutely convex hulls

This section aims at harmonizing the connections of certain operators, Gaussian processes, reproducing kernel Hilbert spaces and absolutely convex hulls as they are employed in the articles Talagrand (1993), Carl (1997), Li and Linde (1999), Li and Linde (2000) and Gao (2004) in one or another form and secondly at giving a precise view of the setting in which our results are developed.

Thereby, we will describe how to transform the purely analytic problem of estimating entropy numbers of $\mathrm{aco}(T)$ to a probabilistic problem, while T is a precompact subset of some Hilbert space H. We may assume the set T to be compact, since passing to the closure does not change entropy behaviour. Furthermore, we assume that $0 \in T$ for technical reasons affecting the supremum of increments in the formulation of Talagrand's lower bound. This does not restrict generality.

We always equip T with the distance induced by the inner product of the underlying Hilbert space H.

Let $l_1(T)$ denote the Banach space of all real valued summable functions over T with the the norm $\|x\|_1 := \sum_{t\in T} |x(t)|$ and define an operator v by

$$v : l_1(T) \to H, \quad v(e_t) := t, \tag{3.18}$$

where $(e_t)_{t\in T}$ denotes the standard canonical base of $l_1(T)$. As we know from Roy (1987), the closed unit ball $B_{l_1(T)}$ is the closed absolutely convex hull of the functions $(e_t)_{t\in T}$. This explains the relation

$$e_n(v) = e_n(\mathrm{aco}(T)).$$

Since $\langle vx, h\rangle = \langle x, v^*(h)\rangle_{l_1,l_\infty}$ for all $x \in l_1(T)$, we have $v^*(h)(t) = \langle t, h\rangle$. Therefrom we can see that the dual operator $v^* : H \to l_\infty(T)$ in fact is a mapping into $C(T)$, the space of all real valued continuous functions over T.

Let $(f_k)_{k\in\mathbb{N}}$ be a complete orthonormal system in H and $(\xi_k)_{k\in\mathbb{N}}$ a sequence of i.i.d. standard normal random variables on some complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We introduce a Gaussian process

$$X_t = \sum_{k\in\mathbb{N}} (v^* f_k)(t)\xi_k = \sum_{k\in\mathbb{N}} \langle t, f_k\rangle \xi_k,\ t \in T. \tag{3.19}$$

The sum in (3.19) converges a.s. for each $t \in T$. To achieve a.s. convergence uniformly in t, it is sufficient to assume Dudley's integral to be finite, i.e.,

$$\int_0^\infty \sqrt{\log N(T, \varepsilon)} d\varepsilon < +\infty, \tag{3.20}$$

as we will recognize in the following. For this purpose, we state a powerful tool in the formulation of Kwapień and Woyczyński(1992).

Theorem 16 (Itô-Nisio)**.** *Let $\eta_1, \eta_2, \ldots$ be a sequence of independent random variables with values in a separable Banach space E. Then the following conditions are equivalent:*

(i). $\sum_{k\in\mathbb{N}} \eta_k$ *converges a.s..*

(ii). $\sum_{k\in\mathbb{N}} \eta_k$ *converges in probability.*

If additionally the random variables $\eta_1, \eta_2, \ldots$ are symmetric, then conditions (i) and (ii) are eqivalent to condition (iii).

(iii) There is a random variable θ with values in E and a separating family $D \subseteq E^*$ such that, for each $a \in D$ the series $\sum_{k\in\mathbb{N}} \langle \eta_k, a\rangle_{E,E^*}$ converges to $\langle \theta, a\rangle_{E,E^*}$ a.s.

Lemma 17. *Let $u : H \to C(T)$ be a bounded linear operator. Then the series*

$$\sum_{k\in\mathbb{N}} (uf_k)(t)\xi_k$$

converges for each $t \in T$ a.s., i.e., the set of convergence may depend on t.

Proof. By Parseval's identity we have

$$\sum_{k\in\mathbb{N}} |(uf_k)(t)|^2 = \sum_{k\in\mathbb{N}} |\langle u^*\delta_t, f_k\rangle|^2 = \|u^*\delta_t\|^2 < +\infty.$$

So, for each $\varepsilon > 0$ and all $m > n > n_0$ we have

$$\mathbb{E}|\sum_{k=1}^{m} (uf_k)(t)\xi_k - \sum_{k=1}^{n} (uf_k)(t)\xi_k|^2 = \sum_{k=n+1}^{m} |(uf_k)(t)|^2 < \varepsilon.$$

Therefore, the sequence of partial sums is Cauchy and hence converges in L^2. By Tschebyschev's inequality we have convergence in probability and applying the Itô-Nisio-Theorem gives a.s. convergence. $\square$

As announced, we can even go further than Lemma 17.

Lemma 18. *Let $u : H \to C(T)$ be a bounded linear operator. If Dudley's integral converges, i.e. $\int_0^\infty \sqrt{\log N(T,\varepsilon)}d\varepsilon < \infty$, the series*

$$\sum_{k\in\mathbb{N}} (uf_k)(\cdot)\xi_k$$

converges in $C(T)$ a.s.

Proof. The finiteness of Dudley's integral implies that there is a continuous modification $Y := (Y_t)_{t\in T}$ of the process $X_t := \sum_{k\in\mathbb{N}}(uf_k)(t)\xi_k,\ t \in T$, i.e. almost all paths of Y are continuous and $\mathbb{P}[X_t = Y_t] = 1$ for all $t \in T$, see Theorem 11.17 in Ledoux and Talagrand (1991).
According to Proposition 9, the process Y may be interpreted as a Gaussian random variable with values in $C(T)$. The relation

$$\sum_{k\in\mathbb{N}} \langle u(f_k), \delta_t\rangle_{C(T),M(T)} = X_t = Y_t = \langle Y, \delta_t\rangle_{C(T),M(T)} \ a.s.$$

is valid, whereas the family of point measures $\{\delta_t | t \in T\} \subseteq M(T)$ is separating for $C(T)$. Applying Itô-Nisio-Theorem yields the result. □

We therefore know that $X = (X_t)_{t\in T}$ defined in (3.19) under (3.20) is a Gaussian process with paths in $C(T)$ a.s. By Proposition 9 we know that X is a Gaussian random variable with values in $C(T)$ a.s. The equation $e_n(v^*) = e_n(K)$, where K denotes the unit ball of rkhs of X as usual, is valid by Proposition 11, while it is easy to compute

$$\int_\Omega a(X)b(X)d\mathbb{P} = \langle v^{**}a, v^{**}b\rangle.$$

Since we also know that $e_n(\mathrm{aco}(T)) = e_n(v)$ it is sufficient to state a connection between $e_n(v)$ and $e_n(v^*)$ in order to connect $e_n(K)$ and $e_n(\mathrm{aco}(T))$, using a deep result of Artstein, Milman, and Szarek (2004):

Theorem 19. *Let $v : E \to F$ be a linear operator between Banach spaces E and F, whereas at least one of the spaces is a Hilbert space. Then there are constants $a, b \geq 1$ independent of v, E, F such that*

$$e_{bn}(v^*) \leq a e_n(v), \quad n \in \mathbb{N}. \tag{3.21}$$

Corollary 20. *In the setting of the preceding theorem, we have*

$$e_{bn}(v) \leq 2a e_n(v^*).$$

Proof. It seems appropriate to pass to the bidual operator $v^{**} : E^{**} \to F^{**}$. Note that there is a canonical factorization in the following way:

$$\begin{tikzcd} E \arrow[r, "v"] \arrow[d, "J_E"'] & F \arrow[d, "J_F"] \\ E^{**} \arrow[r, "v^{**}"'] & F^{**} \end{tikzcd}$$

where the usual embeddings J_E, J_F are metric injections. This yields

$$e_n(v) \leq 2e_n(J_F v) = 2e_n(v^{**} J_E) \leq 2e_n(v^{**}). \tag{3.22}$$

Now, (3.21) and (3.22) imply

$$\frac{1}{2} e_{bn}(v) \leq e_{bn}(v^{**}) \leq a e_n(v^*)$$

and we are done. □

We concentrate the preceding considerations in the next proposition, which is essential for this work.

Proposition 21. *Let T be a compact subset of some Hilbert space fulfilling condition* (3.20)*. Then there are constants $a, b \geq 1$ and a Gaussian random variable X, see* (3.19)*, with values in $C(T)$ in such a way that*

$$e_{b^2 n}(\operatorname{aco}(T)) \leq a e_{bn}(K) \leq a^2 e_n(\operatorname{aco}(T)) \text{ and}$$

$$\log N(\operatorname{aco}(T), \varepsilon) \leq 3b \log N(K, \frac{\varepsilon}{2a}) \leq 9b^2 \log N(\operatorname{aco}(T), \frac{\varepsilon}{4a^2}) \tag{3.23}$$

for small ε, where K denotes the unit ball of the rkhs of X.

Proof. On the one hand, we know that $e_n(v) = e_n(\operatorname{aco}(T))$. On the other hand we have $e_n(v^* : H \to C(T)) = e_n(K)$. Theorem 19 and Corollary 20 relate $e_n(v)$ and $e_n(v^* : H \to l_\infty(T))$. Moreover, the inequality

$$e_n(v^* : H \to l_\infty(T)) \leq e_n(v^* : H \to C(T)) \leq 2e_n(v^* : H \to l_\infty(T))$$

is valid. Hence, there are constants $a, b \geq 1$ so that

$$e_{bn}(K) \leq a e_n(\operatorname{aco}(T)) \tag{3.24}$$

as well as

$$e_{bn}(\mathrm{aco}(T)) \leq a e_n(K). \tag{3.25}$$

It remains to derive (3.23), where we lean onto the inequality (3.24). Let $0 < \varepsilon < e_1(\mathrm{aco}(T))$. This is sufficient for $\frac{\varepsilon}{a} < e_1(\mathrm{aco}(T))$ which yields $N(\frac{\varepsilon}{a}, \mathrm{aco}(T)) > 1$ or equivalently $\log_2 N(\mathrm{aco}(T), \frac{\varepsilon}{a}) \geq 1$. As a consequence, there must exist some $n \geq 2$ with

$$n - 1 \leq \log_2 N(\mathrm{aco}(T), \frac{\varepsilon}{a}) \leq n \tag{3.26}$$

In addition, (3.24) may be written for $n+1$ instead of n as

$$\inf\{\delta > 0 : \log_2 N(K, \delta) \leq (n+1)b - 1\} \leq \inf\{\delta > 0 : \log_2 N(\mathrm{aco}(T), \frac{\delta}{a}) \leq n\}$$

which implies that there is an $\varepsilon_0 \leq 2\varepsilon$ so that $\log_2 N(K, \varepsilon_0) \leq (n+1)b - 1$. From this and (3.26) it follows that (we use $(n+1)b - 1 \leq 3b(n-1)$) for all $n \geq 2$)

$$\frac{1}{3b} \log_2 N(K, 2\varepsilon) \leq \frac{1}{3b} \log_2 N(K, \varepsilon_0) \leq n - 1 \leq \log_2 N(\mathrm{aco}(T), \frac{\varepsilon}{a}),$$

hence

$$\log N(K, 2\varepsilon) \leq 3b \log N(\mathrm{aco}(T), \frac{\varepsilon}{a}).$$

Analogously, one proves

$$\log N(\mathrm{aco}(T), 2\varepsilon) \leq 3b \log N(K, \frac{\varepsilon}{a}),$$

based on (3.25). $\square$

In addition, we show an approach, that will turn out to be usefull in the setting of Proposition 11, in particular in Section 5.1 and is in some sense an extension of Proposition 21.

Proposition 22. *Let $(X_t)_{t \in T}$ be a Gaussian process with paths in $C(T)$ a.s. and $u : H \to C(T)$ be an operator defined on some Hilbert space H in such a way that*

$$\mathbb{E} X_t X_s = \langle u^* \delta_t, u^* \delta_s \rangle_H. \tag{3.27}$$

Then there are constants $a, b \geq 1$ so that

$$e_{b^2 n}(K) \leq 2a e_{bn}(\mathrm{aco}(\{u^* \delta_t : t \in T\})) \leq 2a^2 e_n(K) \text{ and}$$

$$\log N(K,\varepsilon) \leq 3b \log N(\mathrm{aco}(\{u^*\delta_t : t \in T\}), \frac{\varepsilon}{2a})$$
$$\leq 9b^2 \log N(K, \frac{\varepsilon}{8a^2}),$$

respectively, where K denotes the unit ball of the rkhs of X.

Proof. By (3.7), we have $u^*(B_{M(T)}) \supseteq \mathrm{aco}(\{u^*\delta_t : \ t \in T\})$ and hence $e_n(u^*) \geq e_n(\mathrm{aco}(\{u^*\delta_t : \ t \in T\}))$. For an inverse inequality look at

$$u^*(B_{M(T)}) = u^* \left(\overline{aco\{\delta_t : \ t \in T\}}^{wk*} \right) \subseteq \overline{u^*(aco\{\delta_t : \ t \in T\})}^{wk}$$
$$= \overline{u^*(aco\{\delta_t : \ t \in T\})}^{\|\cdot\|},$$

where we used continuity of the operator $u^* : (M(T), wk^*) \to (H, weak)$, which can be seen from Bourbaki (1987), IV, § 1.3, Corollary from Proposition 6. We further used the fact that weak closure of a convex set and its closure in the norm topology coincide in each locally convex space, see Rudin (1973), Theorem 3.12. Consequently, we have

$$e_n(u^*) = e_n(\mathrm{aco}(\{u^*\delta_t : t \in T\})).$$

Equation (3.27) implies

$$\int a(X)b(X)d\mathbb{P} = \langle u^*a, u^*b\rangle_H \tag{3.28}$$

for all $a, b \in E^*$ which can be seen with the reasoning analogue to the proof of Proposition 9. Hence, Proposition 11 applies yielding $e_n(u) = e_n(K)$. □

4. ENTROPY OF CONVEX HULLS

This chapter is devoted to finding general upper bounds for the covering numbers

$$N(\mathrm{aco}(T), \varepsilon)$$

of absolutely convex hulls of subsets T in Hilbert space given upper bounds for the covering numbers $N(T, \varepsilon)$. In section 4.1 we show a completely probabilistic approach to the problem. After some preparations concerning Kuelbs-Li inequalities we prove Theorem 2, parts (i)–(iv) via small deviation estimates, which is possible because of converging Dudley integral. Cf. also Gao (2004), who employs Kuelbs-Li and Kathri-Šidák inequality under a more restrictive condition than converging Dudley integral and does only get the correct bounds in cases (ii) and (iii). We point out that even in cases with not necessarily finite Dudley integral, probabilistic methods do not reach the end of the line: The proof of the remaining cases is due to Lifshits (2009) who noticed that estimates (v) to (vii) follow from Remark 5.3 in Carl, Kyrezi, and Pajor (1999) and gave a proof of the latter cited below. Although in these cases Dudley integral of the original set diverges, he observed that the Dudley integral of the corresponding $\varepsilon-$net is finite, employing Sudakov minoration leads to an upper bound for the entropy of convex hull.
In section 4.2, we give a direct proof of Theorem 2, part (v) as a supplement. These concepts will be accompanied by an access originating in the theory of majorizing measures handling parts (iv) to (vii) of Theorem 2.

4.1 A probabilistic approach

We want to employ probabilistic techniques for Theorem 2 as described in the introduction. For this, we work out one of the Kuelbs-Li inequalities in a convenient form.

4.1.1 Kuelbs-Li inequality for different cases

First of all, let us state a useful consequence of inequality (3.16). For this, we assume $\phi(\varepsilon) \preceq f(\varepsilon)$ for some function $f : (0,\infty) \to (0,\infty)$. Setting $\lambda := 2\sqrt{f(\varepsilon)}$ in (3.16) implies $\log N(K, \varepsilon/\sqrt{f(\varepsilon)}) \leq 2f(\varepsilon) + cf(\varepsilon)$ and hence

$$\log N(K, \varepsilon/\sqrt{f(\varepsilon)}) \preceq f(\varepsilon), \tag{4.1}$$

see Li and Linde (1999). Next we apply (4.1) for two estimates, where the small ball function ϕ has super regular behaviour.

Proposition 23. *Assume for* $0 < \alpha < 2$ *and* $\beta \in \mathbb{R}$ *that*

$$\log \phi(\varepsilon) \preceq \varepsilon^{-\frac{2\alpha}{2-\alpha}} |\log \varepsilon|^{-\frac{2\beta}{2-\alpha}}, \tag{4.2}$$

then

$$\log N(K, \varepsilon) \preceq \varepsilon^{-2} |\log \varepsilon|^{1-\frac{2}{\alpha}} (\log|\log \varepsilon|)^{-2\frac{\beta}{\alpha}}.$$

Proof. From (4.2) we know that there must be a constant $c > 0$ so that

$$\phi(\varepsilon) \leq \exp[c\varepsilon^{-\frac{2\alpha}{2-\alpha}} |\log \varepsilon|^{-\frac{2\beta}{2-\alpha}}] =: f(\varepsilon).$$

Regardless of $\rho, \sigma, c > 0$ for all $0 < \varepsilon < \varepsilon(\rho, \sigma, c)$ we have

$$\varepsilon^{-\rho} \leq \exp[c\varepsilon^{-\sigma}].$$

This yields

$$\begin{aligned}
&\limsup_{\varepsilon \downarrow 0} \varepsilon^2 |\log \varepsilon^{-2} f(\varepsilon)|^{\frac{2}{\alpha}-1} (\log|\log \varepsilon^{-2} f(\varepsilon)|)^{\frac{2\beta}{\alpha}} \\
&\leq c\varepsilon^2 (\varepsilon^{-\frac{2\alpha}{2-\alpha}} |\log \varepsilon|^{-\frac{2\beta}{2-\alpha}})^{\frac{2}{\alpha}-1} |\log(\varepsilon^{-\frac{2\alpha}{2-\alpha}} |\log \varepsilon|^{-\frac{2\beta}{2-\alpha}})|^{2\frac{\beta}{\alpha}} \\
&< +\infty
\end{aligned}$$

and further

$$\limsup_{\varepsilon \downarrow 0} \frac{\log N(K, \varepsilon/\sqrt{f(\varepsilon)})}{f(\varepsilon)} \varepsilon^2 |\log \varepsilon^{-2} f(\varepsilon)|^{\frac{2}{\alpha}-1} (\log|\log \varepsilon^{-2} f(\varepsilon)|)^{\frac{2\beta}{\alpha}} < +\infty.$$

We set $\delta := \varepsilon/\sqrt{f(\varepsilon)}$ and gain

$$\log N(K, \delta) \preceq \delta^{-2} |\log \delta|^{1-\frac{2}{\alpha}} (\log|\log \delta|)^{-\frac{2\beta}{\alpha}}.$$

$\square$

Proposition 24. *If* $\beta > 2$ *and* $\log\log\phi(\varepsilon) \preceq \varepsilon^{-\frac{2}{\beta-2}}$, *then*

$$\log N(K,\varepsilon) \preceq \varepsilon^{-2}(\log|\log\varepsilon|)^{2-\beta}.$$

Proof. By assumption, we know that there must be a constant $c > 0$ so that $\phi(\varepsilon) \leq \exp[\exp[c\varepsilon^{-\frac{2}{\beta-2}}]]$. In particular, in the following we consider $f(\varepsilon) := \exp[\exp[c\varepsilon^{-\frac{2}{\beta-2}}]]$. As an intermediate step, we want to verify that

$$\limsup_{\varepsilon\downarrow 0} \varepsilon^2 \left(\log\log\varepsilon^{-2} f(\varepsilon)\right)^{\beta-2} < +\infty.$$

For this purpose note that regardless of $\rho, \sigma, c > 0$ for all $0 < \varepsilon < \varepsilon(\rho,\sigma,c)$ we have

$$\varepsilon^{-\rho} \leq \exp[\exp[c\varepsilon^{-\sigma}]].$$

This yields

$$\begin{aligned}
\limsup_{\varepsilon\downarrow 0} \varepsilon^2 \left(\log\log\varepsilon^{-2} f(\varepsilon)\right)^{\beta-2} &\leq \limsup_{\varepsilon\downarrow 0} \varepsilon^2 \left(\log\log\exp[2\exp[c\varepsilon^{-\frac{2}{\beta-2}}]]\right)^{\beta-2} \\
&\leq \limsup_{\varepsilon\downarrow 0} \varepsilon^2 \left(\log 2 + c\varepsilon^{-\frac{2}{\beta-2}}\right)^{\beta-2} \\
&< +\infty,
\end{aligned}$$

which leads to

$$\begin{aligned}
&\limsup_{\varepsilon\downarrow 0} \frac{\log N(K,\varepsilon/\sqrt{f(\varepsilon)})}{f(\varepsilon)} \varepsilon^2 \left(\log\log\varepsilon^{-2} f(\varepsilon)\right)^{\beta-2} \\
&\leq \limsup_{\varepsilon\downarrow 0} \frac{\log N(K,\varepsilon/\sqrt{f(\varepsilon)})}{f(\varepsilon)} \limsup_{\varepsilon\downarrow 0} \varepsilon^2 \left(\log\log\varepsilon^{-2} f(\varepsilon)\right)^{\beta-2} \\
&\leq +\infty,
\end{aligned}$$

so we have

$$\log N(K,\varepsilon/\sqrt{f(\varepsilon)}) \preceq f(\varepsilon)\varepsilon^{-2} \left(\log\log\varepsilon^{-2} f(\varepsilon)\right)^{2-\beta}.$$

If we set $\delta := \varepsilon/\sqrt{f(\varepsilon)}$ we get

$$\log N(K,\delta) \preceq \delta^{-2}(\log\log\delta^{-2})^{2-\beta} \approx \delta^{-2}(\log\log\delta^{-1})^{2-\beta}$$

as desired. $\square$

4.1.2 Proof of Theorem 2 parts (i) – (iv)

The compact set T defines a Gaussian process X via (3.19). The proof is split. So, first we cover estimates (i) $-$ (iv), in the next sections we pay attention to estimates (v) $-$ (vii). This division is intrinsic for the probabilistic approach: In cases (i) $-$ (iv) Dudley integral converges and hence X has continuous sample paths a.s. This, in particular, implies that the small ball function $\phi(\varepsilon)$ is finite for all $\varepsilon > 0$, which need not be true if condition (3.20) is not fulfilled.

Proof of (i)– (iv). The proofs of (i) $-$ (iv) share the same argument: We start with an upper bound for $N(T,\varepsilon)$, apply the Talagrand lower bound to find an upper bound for $\phi(\varepsilon)$ and use Kuelbs-Li inequality (3.16) to find an upper bound for $\log N(K,\varepsilon)$. We can apply Proposition 21. Hence, an upper bound of $\log N(K,\varepsilon)$ is also an upper bound for $\log N(\mathrm{aco}(T),\varepsilon)$, where K denotes the unit ball of the rkHs of X.
Let us begin with (i). By Theorem 13, the condition $N(T,\varepsilon) \preceq |\log \varepsilon|^{\beta}$ implies

$$\phi(\varepsilon) \preceq |\log \varepsilon|^{\beta+1}.$$

Further, this and Corollary 2.2. in Aurzada et al. (2009) imply

$$\log N(K,\varepsilon) \preceq |\log \varepsilon|^{\beta+1}.$$

We go on with (ii). Here, we apply Theorem 12. The relation $N(T,\varepsilon) \preceq \varepsilon^{-\alpha}|\log \varepsilon|^{-\beta}$ leads to

$$\phi(\varepsilon) \preceq \varepsilon^{-\alpha}|\log \varepsilon|^{-\beta}.$$

We use this relation combined with Theorem 1 from Li and Linde (1999) to achieve

$$\log N(K,\varepsilon) \preceq \varepsilon^{-\frac{2\alpha}{2+\alpha}}|\log \varepsilon|^{-\frac{2\beta}{2+\alpha}}.$$

In order to verify (iii), we apply Theorem 14 and obtain from $\log N(T,\varepsilon) \preceq \varepsilon^{-\alpha}|\log \varepsilon|^{-\beta}$ that

$$\log \phi(\varepsilon) \preceq \varepsilon^{-\frac{2\alpha}{2-\alpha}}|\log \varepsilon|^{-\frac{2\beta}{2-\alpha}}. \tag{4.3}$$

Now Proposition 23 implies with (4.3) that

$$\log N(K,\varepsilon) \preceq \varepsilon^{-2}|\log \varepsilon|^{1-\frac{2}{\alpha}}(\log|\log \varepsilon|)^{-2\frac{\beta}{\alpha}}.$$

To check (iv) we again apply Theorem 14, which tells us that $\log N(T,\varepsilon) \preceq \varepsilon^{-2}|\log\varepsilon|^{-\beta}$ implies

$$\log\log\phi(\varepsilon) \preceq \varepsilon^{-\frac{2}{\beta-2}}.$$

By Proposition 24 this yields the relation

$$\log N(K,\varepsilon) \preceq \varepsilon^{-2}(\log|\log\varepsilon|)^{2-\beta}.$$

□

4.1.3 Proof of Theorem 2 parts (v) – (vii)

As it will be realized below, assertions (v)–(vii) may be derived by calculus from the following proposition, whose result is already contained in Carl, Kyrezi, and Pajor (1999) in Remark 5.3. The following probabilistic proof of it is due to Lifshits (2009).

Proposition 25. *For any compact set T in Hilbert space we have*

$$\log N(\mathrm{aco}(T), 2\varepsilon) \preceq \left(\frac{1}{\varepsilon}\int_{\varepsilon/2}^{\infty}\sqrt{\log N(T,u)}du\right)^2.$$

Proof. Let S be a minimal $\varepsilon-$net of T. Then we have

$$\log N(S,u) \le \min\{\log N(T,u), \log|S|\} \le \min\{\log N(T,u), \log N(T,\varepsilon)\}.$$

Denote by $(X_s)_{s\in S}$ the Gaussian process, which is associated to S and defined by (3.19). Analogously, let $(X_s^{\mathrm{aco}})_{s\in\mathrm{aco}(S)}$ be the Gaussian process associated to the absolutely convex hull of S. By Dudley's Theorem, see Theorem 11.17 in Ledoux and Talagrand (1991), we have

$$\begin{aligned}\mathbb{E}\sup_{s\in S}|X_s| &\le 48\int_0^{\infty}\sqrt{\log N(S,u)}du \\ &\le 48\left(\varepsilon\sqrt{\log N(T,\varepsilon)} + \int_{\varepsilon}^{\infty}\sqrt{\log N(T,u)}du\right). \end{aligned} \tag{4.4}$$

We observe that

$$\sup_{s\in S}|X_s| = \sup_{s\in\mathrm{aco}(S)}|X_s^{\mathrm{aco}}|. \tag{4.5}$$

In the next step we apply Sudakov minoration, see Theorem 3.18 in Ledoux and Talagrand (1991), which yields for any $\varepsilon > 0$

$$\mathbb{E}\sup_{s\in\mathrm{aco}(S)}|X_s^{\mathrm{aco}}| \ge c\,\varepsilon\sqrt{\log N(\mathrm{aco}(S),\varepsilon)}, \tag{4.6}$$

whereas $c > 0$ is independent of ε. Combining (4.4), (4.5) and (4.6) leads to

$$\log N(\operatorname{aco}(S), \varepsilon) \leq c \left(\sqrt{\log N(T, \varepsilon)} + \frac{1}{\varepsilon} \int_{\varepsilon}^{\infty} \sqrt{\log N(T, u)} du \right)^2 .$$

Note that

$$\sqrt{\log N(T, \varepsilon)} \leq \frac{2}{\varepsilon} \int_{\varepsilon/2}^{\varepsilon} \sqrt{\log N(T, u)} du$$

and hence

$$\log N(\operatorname{aco}(S), \varepsilon) \leq c \left(\frac{1}{\varepsilon} \int_{\varepsilon/2}^{\infty} \sqrt{\log N(T, u)} du \right)^2 .$$

Moreover, we observe that $\operatorname{aco}(T)$ belongs to the $\varepsilon-$neighbourhood of $\operatorname{aco}(S)$: For each $t_i \in T$ there is an $s_i \in S$ so that $\|t_i - s_i\| \leq \varepsilon$ which implies for real numbers $\lambda_1, ..., \lambda_n$ with $\sum_{i=1}^n |\lambda_i| = 1$

$$\| \sum_{i=1}^n \lambda_i t_i - \sum_{i=1}^n \lambda_i s_i \| \leq \sum_{i=1}^n |\lambda_i| \|t_i - s_i\| \leq \varepsilon.$$

Finally, it follows by triangle inequality that

$$\log N(\operatorname{aco}(T), 2\varepsilon) \leq \log N(\operatorname{aco}(S), \varepsilon)$$

and we get the result. □

Proof of Theorem 2, parts (v)– (vii). We assume w.l.o.g. $\varepsilon < \operatorname{diam}(T) < 1$. For proving (v), we apply Proposition 25 which yields

$$\begin{aligned} \log N(\operatorname{aco}(T), 2\varepsilon) &\leq c_1 \left(\frac{1}{\varepsilon} \int_{\varepsilon/2}^{\operatorname{diam}(T)} u^{-1} (-\log u)^{-1} du \right)^2 \\ &= c_1 \varepsilon^{-2} \left(-\log(-\log u)|_{\varepsilon/2}^{\operatorname{diam}(T)} \right)^2 \\ &\leq c_2 \varepsilon^{-2} \left(\log |\log \varepsilon| \right)^2 \end{aligned} \tag{4.7}$$

as it is the assertion.
To show parts (vi) and (vii) we proceed similarly. The only difference is to adjust the integral estimate in (4.7) to

$$\int_{\varepsilon/2}^{\operatorname{diam}(T)} u^{-1} (-\log u)^{-\beta/2} du = \frac{1}{1 - \beta/2} \left(|\log \varepsilon|^{1-\beta/2} - |\log \operatorname{diam}(T)|^{1-\beta/2} \right)$$

and

$$\int_{\varepsilon/2}^{\mathrm{diam}(T)} u^{-\alpha/2} du = \frac{1}{\alpha/2 - 1} \left((\varepsilon/2)^{1-\alpha/2} - (\mathrm{diam}(T))^{1-\alpha/2} \right),$$

respectively. □

4.2 A direct proof of Theorem 2 part (v)

As a supplement, we add a direct proof of (v). We go along the lines of proof of Proposition 2.1, Creutzig and Steinwart (2002), simplify them for Hilbert space setting and adjust them to the case of $\beta = 2$, where originally, the ideas have been applied to the case of $\beta < 2$.

Alternative proof of (v). Since we are only interested in asymptotics for $\varepsilon \to 0$, we restrict ourselves to the case $\varepsilon < 1/2$. Let $c_0 > 0$ denote a constant so that

$$\log N(T, \varepsilon) \leq c_0 \varepsilon^{-2} |\log \varepsilon|^{-2}.$$

For short we write $f(\varepsilon) = \varepsilon^{-2} |\log \varepsilon|^{-2}$. Let $\varepsilon_0 = \gamma 2^{-n}$ for fixed $\gamma \in (1/2, 1]$ and $n \geq 1$.

The definition of $\log N(T, \varepsilon)$ yields that there must exist $\gamma 2^{-k}$-nets N_k with cardinality $|N_k| \leq \exp[c_0 f(\gamma 2^{-k})]$, $k = 1, ..., n$.

We define sets $D_1 := N_1$ and

$$D_k := \{z \in N_k - N_{k-1} : \|z\| \leq \gamma 2^{-k+1}\}, \; k = 2, \dots, n.$$

Note that we have $|D_k| \leq |N_k||N_{k-1}| \leq \exp[2c_0 f(\gamma 2^{-k})]$. For $D_k^{'} := D_k - D_k \cup \{0\}$ it holds that

$$|D_k^{'}| \leq 3|D_k| \leq 3 \exp[2c_0 f(\gamma 2^{-k})].$$

There is a constant $c_1 > 0$ such that $\log 3 + 2c_0 f(\gamma 2^{-k}) \leq c_1 f(\gamma 2^{-k})$, hence

$$|D_k^{'}| \leq \exp[c_1 f(\gamma 2^{-k})], \; k = 1, \dots, n.$$

In a next step, we define $C_k := \mathrm{co}(D_k^{'}) = \mathrm{aco}(D_k)$ and $E_n := \sum_{k=1}^{n} C_k$. For $k \geq 2$ and $t_k \in N_k$ there is a $t_{k-1} \in N_{k-1}$ so that $t_k - t_{k-1} \in D_k$. We apply this fact beginning with $t_n \in N_n$. For these $t_1, \cdots, t_n$ we have $t_n = t_n - t_{n-1} + t_{n+1} - \ldots + t_2 - t_1 + t_1 \in \sum_{k=1}^{n} D_k$. So we can conclude that $N_n \subseteq E_n$. That means, E_n is an ε_0-net for T and – since it is absolutely convex – an ε_0–net for $\mathrm{aco}(T)$. By triangle inequality, we get

$$\log N(\mathrm{aco}(T), 2\varepsilon_0) \leq \log N(E_n, \varepsilon_0). \tag{4.8}$$

In the following, we will prove that

$$M = \{\sum_{k=1}^{n} \frac{1}{m_k} \sum_{i=1}^{m_k} d_{k,i} : d_{k,i} \in D_k^{'}\} = \sum_{k=1}^{n} \frac{1}{m_k} \sum_{i=1}^{m_k} D_k^{'}$$

is an ε_0–net for E_n using an argument due to Maurey, cf. Pisier (1981). Denote the elements of $D_k^{'} \setminus \{0\}$ by $x_1^k, ..., x_{d_k}^k$. Fix a $z \in E_n$ and write $z = \sum_{k=1}^{n} z_k$ with $z_k \in C_k$. Then each z_k can be represented by

$$z_k = \sum_{i=1}^{d_k} a_{k,i} x_i^k, \text{ where } a_{k,i} \geq 0 \text{ and } \sum_{i=1}^{d_k} a_{k,i} \leq 1.$$

Let us define Z_k to be a random vector with values in $D_k^{'}$ by

$$\mathbb{P}(Z_k = x_i^k) = a_{k,i} \text{ for } i = 1, ..., d_k, \text{ and } \mathbb{P}(Z_k = 0) = 1 - \sum_{i=1}^{d_k} a_{k,i}.$$

We obtain $\mathbb{E}Z_k = z_k$. Moreover, take independent random vectors $Z_{1,1}, ...,$ $Z_{1,m_1}, ..., Z_{n,1}, ..., Z_{n,m_n}$ where $Z_{k,i}$ is distributed like Z_k for $k = 1, ..., n$; $i = 1, ..., m_k$. With $Y_{k,i} := \frac{1}{m_k} Z_{k,i}$ and an inequality of Maurey and Pisier, see equation (3) in Creutzig and Steinwart (2002), with a constant $c_2 > 1$ we get

$$\begin{aligned} \mathbb{E}\left\| \sum_{k=1}^{n} z_k - \sum_{k=1}^{n} \frac{1}{m_k} \sum_{i=1}^{m_k} Z_{k,i} \right\| &= \mathbb{E}\left\| \sum_{k=1}^{n} \sum_{i=1}^{m_k} (\mathbb{E}Y_{k,i} - Y_{k,i}) \right\| \\ &\leq c_2 \left(\sum_{k=1}^{n} \sum_{i=1}^{m_k} \mathbb{E}\|Y_{k,i}\|^2 \right)^{1/2} \qquad (4.9) \\ &\leq c_2 \left(\sum_{k=1}^{n} \frac{1}{m_k} 2^{-2(k-1)} \right)^{1/2}. \end{aligned}$$

Let us specify the integers m_k. Since $\varepsilon_0 \leq 2^{-n}$ we observe

$$\frac{k \log(n+1)}{\varepsilon_0^2 2^{2(k-1)}} \geq 1, \quad k = 1, ..., n.$$

Therefore, for each $k = 1, \ldots, n$ there is an integer m_k fulfilling

$$c_2^2 \frac{3k \log(n+1)}{\varepsilon_0^2 2^{2(k-1)}} \leq m_k \leq 2c_2^2 \frac{3k \log(n+1)}{\varepsilon_0^2 2^{2(k-1)}}.$$

We can continue at (4.9) and state

$$\mathbb{E}\left\| z - \sum_{k=1}^{n} \frac{1}{m_k} \sum_{i=1}^{m_k} Z_{k,i} \right\| \leq \varepsilon_0 \frac{1}{\sqrt{3}} \log^{-1/2}(n+1) \left(\sum_{k=1}^{n} k^{-1} \right)^{1/2} \leq \varepsilon_0,$$

since we have $\sum_{k=1}^{n} k^{-1} \leq 3\log(n+1)$ for $n \geq 1$.
Now $\sum_{k=1}^{n} \frac{1}{m_k} \sum_{i=1}^{m_k} Z_{k,i}$ takes only values in the set M, i.e. there must be an $x \in M$ with $\|z - x\| \leq \varepsilon_0$. So M is an ε_0-net for E_n. It remains to calculate

$$\begin{aligned}
\log |M| \leq \log(\prod_{k=1}^{n} |D_k^{'}|^{m_k}) &\leq c_3 \sum_{k=1}^{n} m_k f(\gamma 2^{-k}) \\
&\leq c_4 \varepsilon_0^{-2} \log(n+1) \sum_{k=1}^{n} k^{-1} \\
&\leq c_5 \varepsilon_0^{-2} \log^2(n+1) \\
&\leq c_6 \varepsilon_0^{-2} \log^2 |\log \varepsilon_0|.
\end{aligned}$$

In connection with (4.8), this yields the assertion. □

4.3 Estimates via majorizing measures in the critical case

In this section, we tend to find upper bounds for the entropy of convex hulls with the help of tools arising from the theory of majorizing measures. We mainly refer to the works Li and Linde (2000), Bühler, Li, and Linde (2001) and Bühler (1999).
Initially, that approach seems promising, particularly since we have knowledge of the tight connection between absolutely convex hulls in Hilbert space and Gaussian processes, while majorizing measures are sensitive enough to give both a necessary **and** a sufficient condition for boundednes of a Gaussian process. Moreover, that method could be succesfully applied to finding the correct order of convex hull entropy of the sequence $(\sigma_n f_n)$ of an orthonormal system $(f_n)_{n\in\mathbb{N}}$ in a Hilbert space weighted by $\sigma_n = (\log n)^{-1/2}(\log\log n)^{-b}$, $b > 0$, as done in Li and Linde (2000), cf. section 5.1 for an application facility.
After proving general upper bounds in the critical case, however, we will make some observations reflecting problems in estimating the correct entropy of convex hulls by that approach as well as try to figure out what is behind it.

4.3.1 Basic constructions and properties

Let us introduce those quantities arising from the theory of majorizing measures which will be of use in the following subsection in the same way as done in Li and Linde (2000), section 3.

A precompact subset T of some Hilbert space shall be given. Let $q \geq 16$ be a fixed integer and $i \in \mathbb{Z}$ be the largest integer fulfilling $1 = N(T, q^{-i})$, which depends on T, at the most.

Let $J \subseteq \{i, i+1, i+2, \cdots\} \subset \mathbb{Z}$ be a finite or infinite interval. We denote by $\{\mathcal{A}_j\}_{j \in J}$ a sequence of finite measurable (w.r.t. $\mathfrak{B}(T)$) partitions of T possessing the following three properties:

(i). $\mathcal{A}_i = \{T\}$ whenever $i \in J$,

(ii). $\mathcal{A}_{j+1}$ is a refinement of $\mathcal{A}_j$ for all $j \in J$ and

(iii). each $A \in \mathcal{A}_j$ fulfills $\mathrm{diam}(A) \leq 2q^{-j}$.

We name a sequence $\mathbf{w} = (w_j)_{j \in J}$ of weight functions $w_j : \mathcal{A}_j \to [0,1]$ **adapted** if

(i). $w_i \equiv 1$, whenever $i \in J$ and

(ii). $\sum_{A \in \mathcal{A}_j} w_j(A) \leq 1$ for all $j \in J$.

Let $J = \{i, i+1, \cdots\}$ and set

$$\Theta_{\mathcal{A},w}(T) := \sup_{t \in T} \sum_{j=i+1}^{\infty} q^{-j} \sqrt{|\log w_j(A_j(t))|}. \tag{4.10}$$

where $A_j(t)$ denotes the unique set in $\mathcal{A}_j$ containing t.
Moreover, we define

$$\Theta(T) := \inf\{\Theta_{\mathcal{A},w}(T) : \ \mathcal{A} = (\mathcal{A}_j)_{j \geq i}, \ \mathbf{w} = (w_j)_{j \in J}\}, \tag{4.11}$$

where the infimum is taken over all sequences of finite partitions $\mathcal{A}$ and adapted sequences of weights $\mathbf{w}$ w.r.t. T. For a set $T \subseteq H$ not being precompact, we set $\Theta(T) = \infty$.
This quantity is intensely interwoven with the question of boundedness or unboundedness of the Gaussian Process (3.19) via the following Theorem, see e.g. Talagrand (1996).

Theorem 26 (Talagrand). *There are universal constants $c_1, c_2 > 0$ so that for any $T \subseteq H$,*

$$c_1 \Theta(T) \leq \mathbb{E} \sup_{t \in T} |X_t| \leq c_2 \Theta(T). \tag{4.12}$$

The inequalities in (4.12) express that $\Theta(T)$ is finite if and only if the corresponding Gaussian process is bounded.
For the sake of estimating the entropy of convex hulls, the construction of $\Theta(T)$ is to be modified in two different ways. The first will turn out to be suitable especially for sets T with $\Theta(T) = \infty$, whereas the second is useful if $\Theta(T) < \infty$.
For the first, we regard finite intervals $J = \{i, i+1, \cdots, N\}$ with $N > i$, so in the following, the sequences of partitions $\mathcal{A}$ and weights $\mathbf{w}$ are finite. According to (4.10) and (4.11), respectively, we set

$$\Theta^N_{\mathcal{A},\mathbf{w}}(T) := \sup_{t \in T} \sum_{j=i+1}^{N} q^{-j} \sqrt{|\log w_j(A_j(t))|}$$

and

$$\Theta^N(T) := \inf\{\Theta^N_{\mathcal{A},w}(T) : \ \mathcal{A} = (\mathcal{A}_j)_{j=i}^N, \ \mathbf{w} = (w_j)_{j=i}^N\}.$$

For the second, let $J = \{M+1, M+2, \cdots\}$, hence we handle sequences $\mathcal{A} = (\mathcal{A}_j)_{j>M}$ as well as $\mathbf{w} = (w_j)_{j>M}$ and set

$$\Delta^M_{\mathcal{A},\mathbf{w}}(T) := \sup_{t \in T} \sum_{j=m+1}^{\infty} q^{-j} \sqrt{|\log w_j(A_j(t))|}$$

and

$$\Delta^M(T) := \inf\{\Delta^M_{\mathcal{A},\mathbf{w}}(T) : \ \mathcal{A} = (\mathcal{A}_j)_{j>M}, \ \mathbf{w} = (w_j)_{j>M}\}.$$

To recognize the tight connection of the above constructions to classical majorizing measures the reader be advised to consult Bühler, Li, and Linde (2001) or Talagrand (1996).

The following two propositions are cited from Li and Linde (2000), being corollaries themselves from Proposition 2.1.3 in Bühler (1999).

Proposition 27. *For any $T \subseteq H$ and $N > i$ we have*

$$c_1 \sup_{i \leq j \leq N-1} q^{-j} \sqrt{\log N(T, q^{-j})} \leq \Theta^N(T) \leq c_2 \int_{q^{-N-1}}^{\infty} \sqrt{\log N(T, \varepsilon)} d\varepsilon. \tag{4.13}$$

Proposition 28. *For any $T \subset H$ and $M > i$ we have*

$$c_1 \sup_{j>M} q^{-j}\sqrt{\log N(T, q^{-j})} \le \Delta^M(T) \le c_2 \int_0^{q^{-M}} \sqrt{\log N(T,\varepsilon)}d\varepsilon. \quad (4.14)$$

4.3.2 Proof of Theorem 2, parts (iv)- (vii)

With the help of the findings introduced above, we can provide upper bounds for the entropy of convex hulls in the critical case.

Proof of Theorem 2, parts (iv) – (vi). We start with proving (vi) and consult Theorem 4.1 of Li and Linde (2000), which yields for all $k \in \mathbb{N}$ the relation

$$\sqrt{k}e_k(\mathrm{aco}(T)) \le c_0 \inf_{N>i}\{\Theta^N(T) + q^{-N}\sqrt{k}\}. \quad (4.15)$$

According to Proposition 28 we have

$$\Theta^N(T) \le c_1 \int_{q^{-N-1}}^{\infty} \sqrt{\log N(T,\varepsilon)}d\varepsilon \le c_2 |\log q^{-N-1}|^{1-\beta/2}.$$

For all $k \ge q^{\max\{0,i\}}$ there is an $N \in \mathbb{N}$ so that $q^N \le k \le q^{N+1}$ and

$$\Theta^N(T) \le c_3 (\log k)^{1-\beta/2},$$

which yields

$$e_k(\mathrm{aco}(T)) \le c_4 k^{-1/2}(\log k)^{1-\beta/2}$$

since $1 - \beta/2 > 0$, leading to $\log N(\mathrm{aco}(T),\varepsilon) \preceq \varepsilon^{-2}|\log\varepsilon|^{2-\beta}$.

The only difference in proving (v) is to adjust (4.3.2) to the other integral estimate

$$\int_{q^{-N-1}}^{\infty} \sqrt{\log N(T,\varepsilon)}d\varepsilon \le c \log|\log q^{-N-1}|.$$

To confirm (iv), we employ Proposition 28 telling that

$$\Delta^M(T) \le c_5 \int_0^{q^{-M}} \sqrt{\log N(T,\varepsilon)}d\varepsilon.$$

Let $M(k)$ denote the maximal $M > i$ for which $q^{-M} \ge \varepsilon_k(T)$. Then we have

$$\Delta^{M(k)} \le c_6 \Delta^{M(k)+1} \le c_7 \int_0^{q^{-M(k)-1}} \sqrt{\log N(T,\varepsilon)}d\varepsilon \le c_8 \int_0^{\varepsilon_k(T)} \sqrt{\log N(T,\varepsilon)}d\varepsilon,$$

where Proposition 1.3.2 from Bühler, Li, and Linde (2001) is used for the first inequality, and that leads to

$$\Delta^{M(k)} \leq c_9 (\log\log k)^{1-\beta/2}.$$

This together with Theorem 4.3 from Li and Linde (2000) implies

$$e_k(\mathrm{aco}(T)) \leq c_{10} k^{-1/2} (\log\log k)^{1-\beta/2},$$

and hence $\log N(\mathrm{aco}(T), \varepsilon) \preceq \varepsilon^{-2} (\log |\log \varepsilon|)^{2-\beta}$ which finishes the proof. $\square$

Let us furthermore treat the super-critical case.

Proof of Theorem 2, part (vii). We contact Theorem 4.1 in Li and Linde (2000). Let $k \geq q^{\alpha \max\{i,0\}}$ be given. There is an $N \in \mathbb{N}$ in such a way that

$$q^{-N-1} \leq k^{-1/\alpha} \leq q^{-N}.$$

By Proposition 27 we can calculate that

$$\Theta^N(T) \leq \int_{q^{-N}}^{\infty} \sqrt{\log N(T, \varepsilon)} d\varepsilon \leq c(q^{-N})^{1-\alpha/2} \leq c(k^{-1/\alpha})^{1-\alpha/2} \leq ck^{-1/\alpha+1/2},$$

which implies together with (4.15) that

$$e_k(\mathrm{aco}(T)) \preceq k^{-1/\alpha}$$

corresponding to the assertion. $\square$

4.3.3 Observations on majorizing measures concerning the size of convex hulls

We want to collect some information about two special sets in some Hilbert space in terms of $\Delta_M(\cdot)$ and compare them with what we know about their size of convex hulls in terms of metric entropy.
We again mainly use the Thesis of Bühler (1999) and the articles Bühler, Li, and Linde (2001) as well as Li and Linde (2000). Let T be a subset of Hilbert space as usual. The quantity $\Delta^M(T)$ corresponds to $\Theta_M^\infty(T)$ in the notation of Bühler (1999).
Let us introduce two sets R and S from l_2. We will see that $\Delta^M(R)$ achieves the upper bound in (4.14), while $\Delta^M(S)$ achieves the lower bound, in each

case up to constants independent of M.
Let $(f_k)_{k\in\mathbb{N}}$ be an orthonormal base of l_2. For $b>0$ we define S by

$$S := \{(\log k+1)^{-1/2}(\log\log k+2)^{-b} f_k :\ k\in\mathbb{N}\}$$

and for $b>1$

$$R := \sum_{k\in\mathbb{N}} \{+k^{-1}(\log k+1)^{-b} f_k, -k^{-1}(\log k+1)^{-b} f_k\}.$$

Proposition 29. *For the above defined sets S,R we have*

$$c_1 \sup_{M<l} q^{-l}\sqrt{\log N(S,q^{-l})} \le \Delta^M(S) \le c_2 \sup_{M<l} q^{-l}\sqrt{\log N(S,q^{-l})} \tag{4.16}$$

as well as

$$c_1 \int_0^{q^{-M}} \sqrt{\log N(R,\varepsilon)}d\varepsilon \le \Delta^M(R) \le c_2 \int_0^{q^{-M}} \sqrt{\log N(R,\varepsilon)}d\varepsilon. \tag{4.17}$$

Proof. We start with proving (4.16) and use the Remark on page 17 of Bühler, Li, and Linde (2001) changing roles of M and N:

$$\Theta_M^N(S) \le c_0 \sup_{M\le l\le N+1} q^{-l}\sqrt{\log N(S,q^{-l})}$$

which yields that

$$c_3\Delta^M(S) = c_3\Theta_M^\infty(S) \le c_4 \sup\{\Theta_M^N(S) :\ N<\infty\} \le c_5 \sup_{l\ge M} q^{-l}\sqrt{\log N(S,q^{-l})}.$$

in connection with Theorem 2.4.3 of Bühler (1999). Applying Proposition 1.3.2 of Bühler, Li, and Linde (2001) for the second inequality we find that

$$\sup_{j>M} q^{-j}\sqrt{\log N(S,q^{-j})} \le \Delta^M(S) \le \Delta^{M-1}(S) \le c_6 \sup_{l>M} q^{-l}\sqrt{\log N(S,q^{-l})}.$$

We turn to (4.17). It suffices to find a lower bound. We only use assertions from Bühler (1999) for the rest of the proof. The result of section 2.2.4 loc. cit. is

$$\Gamma_{N_1}^{N_2}(R) \ge \frac{1}{2}\int_{4q^{-N_2}}^{4q^{-N_1}} \sqrt{\log N(R,\varepsilon)}d\varepsilon.$$

By Theorem 2.4.2 we have ($M\Theta_{N_1}^{N_2}$ is fixed notation of Bühler (1999), this M not has anything to do with the M in Proposition 29)

$$c_7 M\Theta_{N_1}^{N_2}(R) \ge \Gamma_{N_2}^{N_1}(R).$$

By Proposition 2.1.2 we have

$$c_8 \Theta_{N_1}^{N_2}(R) \geq M \Theta_{N_1}^{N_2}(R).$$

Theorem 2.4.3 implies

$$\Theta_{N_1}^{\infty}(R) \geq \sup\{\Theta_{N_1}^{N_2}(R) : N_2 < \infty\}.$$

All in all, these assertions imply

$$\begin{aligned} \Delta^{N_1}(R) &\geq c_9 \sup\{\int_{4q^{-N_2}}^{4q^{-N_1}} \sqrt{\log N(R,\varepsilon)} d\varepsilon : N_2 < \infty\} \\ &\geq c_9 \int_0^{q^{-N_1}} \sqrt{\log N(R,\varepsilon)} d\varepsilon. \end{aligned}$$

□

The quantity $\Delta^M(T)$ is used in Li and Linde (2000) to compute upper bounds for Gelfand widths and entropy numbers of convex hulls, e.g. if $M(k) := \sup\{M > i : \ N(T, q^{-M}) \leq k\}$ by the inequalities

$$\sqrt{m} c_{m+k}(T) \leq \Delta^{M(k)}(T)$$

and if β_k is decreasing with

$$\Delta^{M(k)}(T) \leq \beta_k$$

by the inequality

$$\sqrt{k} e_k(\mathrm{aco}(T)) \leq c\beta_k,$$

see Theorems 4.2 and 4.3 loc.cit. One can compute that for $b > 1$

$$\log N(S,\varepsilon) \approx \log N(R,\varepsilon) \approx \varepsilon^{-2} |\log \varepsilon|^{-2b},$$

while

$$\log N(\mathrm{aco}(S),\varepsilon) \approx \varepsilon^{-2} (\log |\log \varepsilon|)^{-2b}$$

due to Li and Linde (2000), Theorem 5, and

$$\log N(\mathrm{aco}(R),\varepsilon) \approx \varepsilon^{-2} |\log \varepsilon|^{-2b}$$

as shown in proof of Proposition 40. We observe that the set S with $\Delta^M(S)$ reaching lower bound of (4.14) has significantly greater entropy of convex hull than the set R, for which $\Delta^M(R)$ reaches upper bound in (4.14) implying the greatest upper bound possible $e_n(\mathrm{aco}(R)) \preceq n^{-\frac{1}{2}}(\log\log n)^{1-b}$ in view of Theorem 2. At a first glance, this seems surprising since one tended to interprete the entity $\Delta^M(T)$ to be a size measure for $\mathrm{aco}(T)$.

Let $(X_t)_{t\in T}$ be a Gaussian process. We write for short

$$\omega_T(t,\varepsilon) := \mathbb{E} \sup_{\substack{s\in T\\ \|t-s\|\le\varepsilon}} |X_t - X_s|$$

while in the following considerations $\omega_T(t,\varepsilon)$ is regarded for the underlying process defined in (3.19).

Bühler, Li, and Linde (2001) tried to gain a description of $\Delta^M(T)$ in known terms, so to speak clarifying the nature of this entity. From Theorem 1.7.2 in Bühler, Li, and Linde (2001) we know that for $N > i$

$$c_1\Delta^M(T) \le \sup_{t\in T}\omega_T(t,q^{-M}) + q^{-M}\sqrt{\log N(T,q^{-M})} \le c_2\Delta^M(T), \quad (4.18)$$

hence, as stated in the subsequent Remark loc. cit., $\Delta^M(T)$ is a combination of both the "local quantity" $\sup_{t\in T}\omega_T(t,q^{-M})$ and "global quantity" $q^{-M}\sqrt{\log N(T,q^{-M})}$ which should be necessary to appear in general. In the next lemma, however, we recognize that $\Delta^M(T)$ is asymptotically eqivalent only to the local quantity $\sup_{t\in T}\omega_T(t,q^{-M})$. This one is hardly to be recognized as reflecting the geometry of the original set, which is what one could expect as helpful or even necessary in order to estimate the actual size of convex hulls.

Lemma 30. *If we have* $\log N(T,\varepsilon) \approx \varepsilon^{-2}J(1/\varepsilon)$ *with a function* $J(\cdot)$ *slowly varying at infinity, then*

$$\sup_{t\in T}\omega_T(t,\varepsilon) \succeq \varepsilon\sqrt{\log N(T,\varepsilon)}. \quad (4.19)$$

Proof. We suppose

$$\sup_{t\in T}\omega_T(t,\varepsilon) \preceq \varepsilon\sqrt{\log N(T,\varepsilon)}f_0(\varepsilon) \quad (4.20)$$

for a function $f_0 \in o(1)$ while assuming $\log N(T,\varepsilon) \approx \varepsilon^{-2}J(1/\varepsilon)$.

By Sudakov minoration, see Theorem 3.18 in Ledoux and Talagrand (1991), there is a universal constant c_0 in such a way that

$$c_0\delta\sqrt{\log N(B_\varepsilon(t),\delta)} \le \mathbb{E}\sup_{s\in B_\varepsilon(t)} X_s \le \omega_T(t,\varepsilon)$$

and equivalently

$$\log N(B_\varepsilon(t),\delta) \le \frac{\omega_T(t,\varepsilon)^2}{c_0^2\delta^2}$$

for each $t \in T$.
We take into account relation (4.20) , i.e., there is a constant c_1 so that for all ε sufficiently small

$$\sup_{t\in T}\omega_T(t,\varepsilon) \le c_1\varepsilon\sqrt{\log N(T,\varepsilon)}f_0(\varepsilon) =: S(\varepsilon).$$

This enables us to estimate

$$N(T,\delta) \le N(T,\varepsilon)\exp\left[\frac{S(\varepsilon)^2}{c_0^2\delta^2}\right],$$

and taking logarithm yields

$$\begin{aligned}\log N(T,\delta) &\le \log N(T,\varepsilon) + \frac{\varepsilon^2\log N(T,\varepsilon)}{\delta^2}f_1(\varepsilon)\\ &\le \log N(T,\varepsilon)\left(1+f_1(\varepsilon)\frac{\varepsilon^2}{\delta^2}\right)\end{aligned}$$

for a function $f_1 \in o(1)$. We remind of the assumption

$$c_2\varepsilon^{-2}J(1/\varepsilon) \le \log N(T,\varepsilon) \le c_3\varepsilon^{-2}J(1/\varepsilon)$$

for $0 < c_2 \le c_3$ and ε sufficiently small. This leads to

$$\frac{c_2}{c_3}\frac{\delta^{-2}J(1/\delta)}{\varepsilon^{-2}J(1/\varepsilon)} \le \frac{\log N(T,\delta)}{\log N(T,\varepsilon)}.$$

In fact, we have by (30)

$$\frac{\varepsilon^2J(1/\delta)}{\delta^2J(1/\varepsilon)} \le \frac{c_3}{c_2} + f_2(\varepsilon)\frac{\varepsilon^2}{\delta^2},$$

with $f_2 \in o(1)$.
If we set $\delta := \sqrt{\frac{c_2}{c_3}}(1-h)\varepsilon$ for an $0<h<1$, we get

$$\frac{1}{(1-h)^2}\frac{J(1/(\sqrt{\frac{c_2}{c_3}}(1-h)\varepsilon))}{J(1/\varepsilon)} \le 1 + f_3(\varepsilon)\frac{1}{(1-h)^2}$$

for $f_3 \in o(1)$ which means

$$\frac{J(1/(\sqrt{\frac{c_2}{c_3}}(1-h)\varepsilon))}{J(1/\varepsilon)} \leq (1-h)^2 + f_3(\varepsilon).$$

Since J was assumed to be slowly varying, this leads to a contradiction as ε tends to zero. Hence,

$$\sup_{t \in T} \omega_T(t, \varepsilon) \succeq \varepsilon \sqrt{\log N(T, \varepsilon)},$$

must be valid, which is the assertion. □

As a consequence, (4.18) now reads as

$$c_4 \Delta^M(T) \leq \sup_{t \in T} \omega_T(t, q^{-M}) \leq c_5 \Delta^M(T),$$

for all M sufficiently large.

We thank M. A. Lifshits for a discussion, in which the preceding lemma emerged.

5. CRITICALLY LARGE ENTROPY

In this chapter, we collect several results in the range of critically large entropy beginning with section 5.1, where we provide lower estimates for the small deviation function based on a lower bound for the metric entropy of the unit ball of rkHs of the related Grv. These are combined with the upper bounds for the small deviations of Gaussian independent sequences and a Volterra type process computed in subsections 5.1.1 and 5.1.2 to find the correct orders of the corresponding small deviation functions.
In a next step, i.e. section 5.2, we provide some observations concerning Kuelbs-Li inequalities in the critical case. The last step of this work is to discuss Gao set. In particular, we try to find lower bounds in terms of entropy numbers by a technique differing from Gao lemma, see Lemma 2.2 in Creutzig and Steinwart (2002).

5.1 Lower estimates

In this section we want to employ the Kuelbs-Li inequality (3.16) to obtain lower estimates for the small deviations of some Gaussian processes if a lower bound for the entropy of the unit ball K of the reproducing kernel Hilbert space is known.
In Kuelbs and Li (1993) and Li and Linde (1999) this was only done in the non-critical case while underlining that for a Gaussian random variable with values in a separable Banach space

$$\log N(K,\varepsilon) \in o(\varepsilon^{-2})$$

necessarily holds. Of course this is fulfilled if $\log N(K,\varepsilon) \preceq \varepsilon^{-\alpha}$, $0<\alpha<2$, but also in our critical cases considered below.

Lemma 31. *Let* $\log N(K,\varepsilon) \succeq \varepsilon^{-2} J(1/\varepsilon)$ *for* J *slowly varying. Then*

$$|\log \varepsilon/\sqrt{\phi(\varepsilon)}| \approx \log \phi(\varepsilon).$$

Proof. We want to use a discrete version of (4.1), which may be formulated as follows: Let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence of positive numbers converging

monotonously to 0 and let $\phi(\varepsilon_n) \leq c_1\varepsilon_n^{-\alpha}$. This implies $\log N(K,\delta_n) \leq c_2\delta_n^{-\frac{2\alpha}{2+\alpha}}$ for $\alpha > 0$ where $\delta_n := \varepsilon_n/\sqrt{\varepsilon_n^{-\alpha}}$.
We claim that there is an $\varepsilon_0 > 0$, so that we have

$$|\log 1/\varepsilon| \leq \log \phi(\varepsilon)$$

for all $0 < \varepsilon < \varepsilon_0$. Suppose not! Then there would be a sequence (ε_n) converging to 0 and fulfilling $\phi(\varepsilon_n) \leq \varepsilon_n^{-1}$. This would yield $\log N(K,\delta_n) \leq c_3\delta_n^{-2/3}$ for all $n \geq n_0$ and $\delta_n = \varepsilon_n^{3/2}$. This contradicts the assumption. Therefore, the inequalities

$$\frac{1}{2}\log\phi(\varepsilon) \leq |\log\varepsilon - \frac{1}{2}\log\phi(\varepsilon)| \leq (1+\frac{1}{2})\log\phi(\varepsilon).$$

hold true for sufficiently small ε. □

Proposition 32. *Let $b > 0$.*
If $\log N(K,\varepsilon) \succeq \varepsilon^{-2}(\log|\log\varepsilon|)^{-2b}$, then

$$\log\log\phi(\varepsilon) \succeq \varepsilon^{-\frac{1}{b}}. \tag{5.1}$$

If $\log N(K,\varepsilon) \succeq \varepsilon^{-2}|\log\varepsilon|^{-2b}$, then

$$\log\phi(\varepsilon) \succeq \varepsilon^{-\frac{1}{b}}. \tag{5.2}$$

If $\log N(K,\varepsilon) \succeq \varepsilon^{-2}|\log\varepsilon|^{2-2b}$ for $b > 1$, then

$$\log\phi(\varepsilon) \succeq \varepsilon^{-\frac{1}{b-1}} \tag{5.3}$$

If $\log N(K,\varepsilon) \succeq \varepsilon^{-\frac{2}{2b-1}}$ for $b > 1$, then

$$\log\phi(\varepsilon) \succeq \varepsilon^{-\frac{2}{2b-1}}. \tag{5.4}$$

Proof. We use (3.16) to obtain $\phi(\varepsilon) \succeq g(\varepsilon/\sqrt{\phi(\varepsilon)})$ whenever $\log N(K,\varepsilon) \succeq g(\varepsilon)$, see Li and Linde (1999), for some regular varying function $g(\cdot)$. This yields

$$\limsup_{\varepsilon\to 0} \frac{\varepsilon^{-2}\phi(\varepsilon)(\log|\log\varepsilon/\sqrt{\phi(\varepsilon)}|)^{-2b}}{\phi(\varepsilon)} < +\infty. \tag{5.5}$$

We apply Lemma 31 and get

$$\limsup_{\varepsilon\to 0} \frac{\varepsilon^{-2/2b}}{(\log|\log\phi(\varepsilon)|)} < +\infty, \tag{5.6}$$

which is the result for (5.1). The second assertion (5.2) and the fourth one (5.4) are proven analogously; the third assertion (5.3) is a simple reformulation of (5.2) . □

In particular, Proposition 32 applies to sequences of weighted Gaussian random variables. For notational simplicity, let us assume for the weights that $\sigma_1 = \sigma_2 = \sigma_3 = 1$. If σ_n is given in one or another form, this is meant for $n \geq 4$.

Corollary 33. *Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. $\mathcal{N}(0,1)-$random variables. If $\sigma_n = (\log n)^{-1/2}(\log\log n)^{-b}$, where $b > 0$, then*

$$\log\log\left|\log \mathbb{P}\left[\sup_{n\in\mathbb{N}} |\sigma_n \xi_n| \leq \varepsilon\right]\right| \succeq \varepsilon^{-\frac{1}{b}}. \tag{5.7}$$

If $\sigma_n = (\log n)^{-b}$, where $b > \frac{1}{2}$, then

$$\log\left|\log \mathbb{P}\left[\sup_{n\in\mathbb{N}} |\sigma_n \xi_n| \leq \varepsilon\right]\right| \succeq \varepsilon^{-\frac{2}{2b-1}}. \tag{5.8}$$

Proof. Let $(f_n)_{n\in\mathbb{N}}$ be an orthonormal system in l_2, $T := \{\sigma_n f_n :\ n \in \mathbb{N}\}$ and $X = (X_t)_{t\in T}$ be the Gaussian process defined in (3.19). We then have

$$\mathbb{P}\left[\sup_{t\in T} |X_t| \leq \varepsilon\right] = \mathbb{P}\left[\sup_{n\in\mathbb{N}} |\sigma_n \xi_n| \leq \varepsilon\right]. \tag{5.9}$$

For the case of $\sigma_n = (\log n)^{-1/2}(\log\log n)^{-b}$, we employ Theorem 5.1 in Li and Linde (2000), which yields

$$\log N(\mathrm{aco}(\{\sigma_n f_n :\ n \in \mathbb{N}\}), \varepsilon) \approx \varepsilon^{-2}(\log|\log\varepsilon|)^{-2b},$$

hence

$$\log N(K, \varepsilon) \approx \varepsilon^{-2}(\log|\log\varepsilon|)^{-2b}.$$

Applying Proposition 32 yields (5.7). Note that there is no problem in the case of $0 < b \leq 1$ where Dudley's integral does not converge: We know from (4.16) that

$$\Delta^M(\{(\log n)^{-1/2}(\log\log n)^{-b} f_k : k \in \mathbb{N}\}) \to 0$$

as $M \to \infty$ for $b > 0$. This is necessary and sufficient for $(\sigma_n \xi_n)_{n\in\mathbb{N}}$ to have uniformly continuous paths according to Corollary 1.7.3 of Bühler, Li, and Linde (2001).

For the case of $\sigma_n = (\log n)^{-b}$ we get from the proof of Proposition 5.5 in Carl, Kyrezi, and Pajor (1999) that

$$\log N(\mathrm{aco}(\{\sigma_n f_n :\ n \in \mathbb{N}\}), \varepsilon) \approx \varepsilon^{-2}|\log\varepsilon|^{1-2b},$$

which means that

$$\log N(K,\varepsilon) \approx \varepsilon^{-2}|\log\varepsilon|^{1-2b}.$$

Again, by Proposition 32 we get the result (5.8). □

Of course, we are also interested in upper bounds for the small deviations of these sequences leading over to the next section.

5.1.1 Application: Small deviations of a Gaussian sequence

There has been a lot of interest in the small deviations of sequences $(\sigma_n\xi_n)_{n\in\mathbb{N}}$, even for more general cases than $distr(\xi_n) = \mathcal{N}(0,1)$, and norms different from the sup −norm, see Borovkov and Ruzankin (2008) or Aurzada (2006) for further information as well as for concrete results.

However, in both articles the considered sequences of weights $(\sigma_n)_{n\in\mathbb{N}}$ are of polynomially decreasing order. In Aurzada (2008), exponentially decreasing sequences of weighted positive i.i.d. random variables are considered.

Hence, the sequences of Corollary 33 are not covered in these papers. We point out that upper bounds for the small deviations in these two cases have been found by Weber (2010), while establishing a general upper bound for small deviations with the help of majorizing measures, see Remark 35 for more details. The next proposition is joint work with Aurzada (2010).

Proposition 34. *Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. $\mathcal{N}(0,1)-$random variables. If $\sigma_n = \left(\log n(\log\log n)^b\right)^{-1/2}$ with $b>0$, then*

$$\log\log\left|\log\mathbb{P}\left[\sup_{n\in\mathbb{N}}|\sigma_n\xi_n|\le\varepsilon\right]\right| \approx \varepsilon^{-\frac{2}{b}}. \tag{5.10}$$

If $\sigma_n = (\log n)^{-b}$ with $b>\frac{1}{2}$, then

$$\log\left|\log\mathbb{P}\left[\sup_{n\in\mathbb{N}}|\sigma_n\xi_n|\le\varepsilon\right]\right| \approx \varepsilon^{-\frac{2}{2b-1}}. \tag{5.11}$$

Proof. We start with proving the first assertion (5.10). Clearly we have

$$\mathbb{P}\left[\sup_{n\in\mathbb{N}}\sigma_n|\xi_n|\le\varepsilon\right] = \prod_{n:\varepsilon<\sigma_n}\mathbb{P}\left[\sigma_n|\xi|\le\varepsilon\right]\prod_{n:\varepsilon\ge\sigma_n}\mathbb{P}\left[\sigma_n|\xi|\le\varepsilon\right]$$

and thus

$$-\log\mathbb{P}\left[...\right] = -\sum_{n:\varepsilon<\sigma_n}\log\mathbb{P}\left[\sigma_n|\xi|\le\varepsilon\right] - \sum_{n:\varepsilon\ge\sigma_n}\log\mathbb{P}\left[\sigma_n|\xi|\le\varepsilon\right]. \tag{5.12}$$

We will treat both sums separately.
Let us consider the first one. We use for short the notation $f(x) = (\log x)^b$ and have $\varepsilon < \sigma_n$ if and only if $\varepsilon^{-2} > \log n f(\log n)$. This gives $\log n \leq c_1 \varepsilon^{-2} f(\varepsilon^{-2})$ for $\varepsilon > 0$ sufficiently small. Therefore, the number of summands is bounded by $\exp[c_1 \varepsilon^{-2} f(\varepsilon^{-2})]$ and we can estimate the first sum from above by

$$-\log \mathbb{P}\left[|\xi| \leq \varepsilon\right] \cdot \sum_{n:\varepsilon<\sigma_n} 1 \leq |\log \varepsilon| \exp[c_1 \varepsilon^{-2} f(\varepsilon^{-2})]. \tag{5.13}$$

We turn to evaluating the second sum in (5.12) and know that

$$\lim_{x\to 0} \frac{|\log(1-x)|}{x} = 1.$$

Moreover, the sequence $\left(\mathbb{P}\left[|\xi| > \frac{\varepsilon}{\sigma_n}\right]\right)_n$ is bounded away from 1 by $\mathbb{P}\left[|\xi| > 1\right]$. This implies that there is a constant $c_2 > 0$ independent of ε so that

$$\frac{1}{c_2}\mathbb{P}\left[|\xi| > \frac{\varepsilon}{\sigma_n}\right] \leq |\log(1 - \mathbb{P}\left[|\xi| > \frac{\varepsilon}{\sigma_n}\right])| \leq c_2 \mathbb{P}\left[|\xi| > \frac{\varepsilon}{\sigma_n}\right].$$

As a consequence, we have the inequality

$$\frac{1}{c_2} \sum_{n:\varepsilon\geq\sigma_n} \mathbb{P}\left[|\xi| > \frac{\varepsilon}{\sigma_n}\right] \leq \sum_{n:\varepsilon\geq\sigma_n} |\log \mathbb{P}\left[\sigma_n|\xi| \leq \varepsilon\right]| \leq c_2 \sum_{n:\varepsilon\geq\sigma_n} \mathbb{P}\left[|\xi| > \frac{\varepsilon}{\sigma_n}\right].$$

Hereby, it is sufficient to find the asymptotic rate for

$$\sum_{n:\varepsilon\geq\sigma_n} \mathbb{P}\left[|\xi| > \frac{\varepsilon}{\sigma_n}\right]$$

as $\varepsilon \to 0$, in order to find the rate of the second sum in (5.12).
Integration by parts yields

$$\int_t^\infty u \exp[-u^2/2] 1/u \; du = 1/t \exp[-t^2/2] - \int_t^\infty \exp[-u^2/2] 1/u^2 \; du,$$

for $t > 0$ and hence

$$\mathbb{P}\left[|\xi| > t\right] = \mathbb{P}\left[\xi < -t\right] + \mathbb{P}\left[\xi > t\right] \leq 2\exp[-t^2/2], \; t \geq 1.$$

To obtain an upper bound for the second sum in (5.12) it thereby suffices to examine the sum

$$\sum_{n\geq 4} \exp[-\frac{1}{2}(\varepsilon/\sigma_n)^2]. \tag{5.14}$$

Since the summands are comparable in the sense that

$$1 \leq \frac{\exp[-\frac{1}{2}(\varepsilon/\sigma_n)^2]}{\exp[-\frac{1}{2}(\varepsilon/\sigma_{n-1})^2]} \leq c_3,$$

we can estimate the sum in (5.14) by an integral and consider

$$\int_3^\infty \exp[-\frac{1}{2}\varepsilon^2 \log n f(\log n)]dn. \tag{5.15}$$

First, we substitute $x := \log n$, hence $dn = \exp[x]dx$ which yields

$$\int_{\log 3}^\infty \exp[-\frac{1}{2}\varepsilon^2 x f(x) + x]dx. \tag{5.16}$$

For technical reasons, we raise the lower integral bound $\log 3$ to a constant $c(b) > \max\{\exp[\exp[b]], \exp[\exp[1]]\}$ only depending on b. This does not change the asymptotic behaviour of the integral (5.16) as $\varepsilon \to 0$, since

$$\begin{aligned}
&\int_{c(b)}^\infty \exp[-\frac{1}{2}\varepsilon^2 x f(x) + x]dx \\
&\leq (5.16) \\
&\leq (c(b) - \log 3)\exp[c(b)] + \int_{c(b)}^\infty \exp[-\frac{1}{2}\varepsilon^2 x f(x) + x]dx.
\end{aligned}$$

The next step will be to substitute $y := xf(x)$. Note that the function

$$g : [c(b), \infty) \to [c(b)f(c(b)), \infty),\ g(x) := xf(x)$$

is continuous and strictly increasing, hence the inverse function

$$g^{-1} : [c(b)f(c(b)), \infty) \to [c(b), \infty)$$

exists and is continuous and strictly increasing as well. Formal result of the substitution is at once

$$\int_{g(c(b))}^\infty \exp[-\frac{1}{2}\varepsilon^2 y + g^{-1}(y)]\frac{dx}{dy}dy. \tag{5.17}$$

We cannot give an explicit formula for g^{-1}. Similar to Aurzada (2006), p.32, we introduce the function

$$S : [g(c(b)), \infty) \to [g(c(b))f(g(c(b)))^{-1}, \infty),\ S(y) := yf(y)^{-1},$$

which is strictly monotonic, and compute

$$\lim_{y\to\infty}\frac{g^{-1}(y)}{S(y)}=\lim_{x\to\infty}\frac{x}{S(g(x))}=\lim_{x\to\infty}\frac{f(xf(x))}{f(x)}=1.$$

Therefore, there is a constant $c_4>0$ so that

$$\int_{g(c(b))}^{\infty}\exp[-\frac{1}{2}\varepsilon^2 y+\frac{g^{-1}(y)}{S(y)}S(y)]\frac{dx}{dy}dy\le\int_{g(c(b))}^{\infty}\exp[-\frac{1}{2}\varepsilon^2 y+c_4 S(y)]\frac{dx}{dy}dy. \tag{5.18}$$

Moreover, we observe that S is a diffeomorphism. For $B\in\mathfrak{B}([g(c(b)),\infty))$ and λ the Lebsegue measure on this $\sigma-$algebra we have by Lebesgue Substitution Theorem, see Theorem 6.1.6 in Cohn (1980),

$$\lambda(B)=\int_B d\lambda=\int_{S(B)}\frac{dS^{-1}}{d\lambda}d\lambda\circ S\circ S^{-1}=\int_B\frac{dS^{-1}}{d\lambda}\circ S\;d\lambda\circ S. \tag{5.19}$$

It follows that λ is equivalent to $\lambda\circ S$. The same arguments as for (5.19) hold true for $\lambda\circ g^{-1}$ which in turn is eqivalent to λ and $\lambda\circ S$.
For any interval $[a,b]\subseteq[g(c(b)),\infty)$ we have that

$$\lambda\circ S([a,b])=S(b)-S(a)$$

– analogously for g^{-1}. Hereby $\lambda\circ S$ and $\lambda\circ g^{-1}$ are the measures generated by S and g^{-1}, respectively, see Bauer (1990), 6.5 Satz. By equivalence of the measures,

$$\frac{dx}{dy}=\frac{dg^{-1}}{dy}=\frac{dg^{-1}}{dS}\frac{dS}{dy} \tag{5.20}$$

is valid. There is a constant c_5 so that

$$\frac{dg^{-1}}{dS}(y)=\frac{\frac{dg^{-1}}{dy}(y)}{\frac{dS}{dy}(y)}=\frac{1}{g'(g^{-1}(y))S'(y)}\le c_5,\quad y\in[g(c(b)),\infty), \tag{5.21}$$

since $g'(x)(S\circ g)'(x)$ is bounded away from zero on $[c(b),\infty)$, to be more precise

$$\begin{aligned}(S\circ g)'(x)=&\left(\frac{\log x}{\log x+b\log\log x}\right)^b\\&+bx\left(\frac{\log x}{\log x+b\log\log x}\right)^{b-1}\frac{\frac{1}{x}(\log x+b\log\log x)-\log x(\frac{1}{x}+b\frac{1}{\log x}\frac{1}{x})}{(\log x+b\log\log x)^2},\end{aligned}$$

where the first summand is strictly positive on $[c(b), \infty)$, continuous and converges to 1 as $x \to \infty$ and the second summand is also positive for all $x \in [c(b), \infty)$.
Now, we can continue with (5.18) using (5.20), (5.21) and $S^{'}(y) \leq (\log y)^{-b}$ which yields

$$\int_{g(c(b))}^{\infty} \exp[-\frac{1}{2}\varepsilon^2 y + c_1 S(y)] \frac{dx}{dy} dy \leq \int_{g(c(b))}^{\infty} \exp[-\frac{1}{2}\varepsilon^2 y + c_1 S(y)] c_6 (\log y)^{-b} dy$$

We substitute $y = z^{-1}$, so $dy = -z^{-2} dz$ and get

$$\int_0^{g(c(b))^{-1}} \exp[-\frac{1}{2}\varepsilon^2 z^{-1} + c_1 S(z^{-1})] c_6 z^{-2} (\log z^{-1})^{-b} dz. \tag{5.22}$$

of course, by changing the constant c_1 we can estimate

$$(5.22) \leq \int_0^{g(c(b))^{-1}} \exp[-\frac{1}{2}\varepsilon^2 z^{-1} + c_7 S(z^{-1})] dz. \tag{5.23}$$

The following part of the proof is aimed at showing that the integrand in (5.23) has a local maximum in in the intervall $(0, \frac{1}{\exp[\exp[b]]})$. Knowing this will suffice to find further reasonable estimates from above.
Let

$$h(z) := \exp[-\frac{1}{2}\varepsilon^2 z^{-1} + c_7 S(z^{-1})] = \exp[-\frac{1}{2}\varepsilon^2 z^{-1} + c_7 z^{-1} (\log z^{-1})^{-b}].$$

We provide the first derivative of h by

$$\begin{aligned} h'(z) &= \exp[\cdots] \left[\frac{1}{2}\varepsilon^2 z^{-2} - c_7 z^{-2} (\log z^{-1})^{-b} - b c_7 z^{-1} (\log z^{-1}) z (-z^{-2}) \right] \\ &= \exp[\cdots] z^{-2} \left[\frac{1}{2}\varepsilon^2 - c_7 (\log z^{-1})^{-b} + b c_7 (\log z^{-1})^{-b-1} \right] \\ &= \exp[\cdots] z^{-2} \left[\frac{1}{2}\varepsilon^2 - c_7 (\log z^{-1})^{-b} \left(1 - b (\log z^{-1})^{-1}\right) \right]. \end{aligned}$$

Firstly, we see that $h^{'}$ is strictly positive in a (possibly small, depending on the constants) neighbourhood of zero intersected with the positive halfline because

$$\lim_{z \downarrow 0} c_7 (\log z^{-1})^{-b} \left(1 - b(\log z^{-1})^{-1}\right) = 0.$$

Secondly, there must be a $\delta(b) > 0$ so that $h^{'}$ is negative on $(\frac{1}{\exp[\exp[b]]} - \delta(b), \frac{1}{\exp[\exp[b]]})$ for sufficiently small $\varepsilon > 0$ since

$$\lim_{z \uparrow \frac{1}{\exp[\exp[b]]}} \frac{b}{\log \frac{1}{z}} - 1 \leq \frac{b}{\exp[b]} - 1 < 0.$$

Therefore, $h(z)$ has at least one local maximum on $(0, \frac{1}{\exp[\exp[1]]})$ or $(0, \frac{1}{\exp[\exp[b]]})$, respectively, and note that the number of maxima is finite. The largest one among all of them shall be attained, say, in $z(\varepsilon)$, and is a global maximum at the same time, while

$$\begin{aligned} \frac{1}{2}\varepsilon^2 &= c_7(\log z(\varepsilon)^{-1})^{-b} - bc_7(\log z(\varepsilon)^{-1})^{-b-1} \\ &= c_7(\log z(\varepsilon)^{-1})^{-b}(1 - b(\log z(\varepsilon)^{-1})^{-1}). \end{aligned} \tag{5.24}$$

Note that $\frac{1}{g(c(b))} \leq \frac{1}{\exp[\exp[b]]}$. By means of the previous argumentation, we can go on with (5.23) and estimate an upper bound by

$$(5.23) \leq g(c(b))^{-1} \exp[-\frac{1}{2}\varepsilon^2 z(\varepsilon)^{-1} + c_7 z(\varepsilon)^{-1}(\log z(\varepsilon)^{-1})^{-b}]. \tag{5.25}$$

As can be seen from (5.24), $z(\varepsilon)$ necessarily tends to zero as ε tends to zero: Of course, the right hand side in (5.24) would also converge to zero, if $z(\varepsilon)$ would converge to $\exp[-b]$. This cannot happen, we know that $z(\varepsilon) \leq \frac{1}{\exp[\exp[b]]} < \frac{1}{\exp[b]}$. Consequently, there are constants $c_8, c_9 > 0$ so that we get from (5.24) the inequality

$$\frac{1}{2}\varepsilon^2 \leq c_8(\log z(\varepsilon)^{-1})^{-b}$$

as well as from (5.25) the relation

$$\log\log(5.23) \leq c_9 \log z(\varepsilon)^{-1}$$

implying

$$\log\log(5.23) \leq c_{10}\varepsilon^{-\frac{2}{b}}$$

for all ε sufficiently small. This is in fact the result, since the asymptotic rate of the first sum in (5.13) is negligible. Combined with Corollary 33 we get

$$\log\log\phi(\varepsilon) \approx \varepsilon^{-2/b}.$$

All in all we have identified the correct order of the small deviations.

It remains to show assertion (5.11). It is known that

$$e_n(\{\sigma_k f_k : \ k \in \mathbb{N}\}) \approx n^{-b},$$

hence

$$\log N(\mathbb{N}, \varepsilon) \approx \varepsilon^{-1/b}$$

with respect to Dudley metric induced by $(\sigma_n \xi_n)_{n\in\mathbb{N}}$. Theorem 14 gives

$$\log \phi(\varepsilon) \preceq \varepsilon^{-\frac{2}{2b-1}}. \tag{5.26}$$

Corollary 33 leads to

$$\log \phi(\varepsilon) \succeq \varepsilon^{-\frac{2}{2b-1}},$$

which is the assertion combined with (5.26) . □

Remark 35. As announced in the introduction of this section, let us have a look on the results of Weber (2010) concerning small deviations of Gaussian sequences $(\sigma_n \xi_n)_{n\in\mathbb{N}}$.

For the case $\sigma_n := (\log n)^{-b}, \ b > \frac{1}{2}$ the upper bound

$$\log \left| \log \mathbb{P} \left[\sup_{n\in\mathbb{N}} |\sigma_n \xi_n| \leq \varepsilon \right] \right| \preceq \varepsilon^{-\frac{2}{2b-1}} \tag{5.27}$$

is computed, using the general lower bound provided by Theorem 3.1 loc. cit. This coincides with the upper bound due to Theorem 3 in Aurzada and Lifshits (2008).

In fact, it is not surprising that both general theorems lead to the correct order, since the Hilbert space counterpart $\{\sigma_n f_n : n \in N\} \subset l_2$ has maximal metric entropy of convex hull under the regime of $\log N(\{\sigma_n f_n : n \in N\}, \varepsilon) \approx \varepsilon^{-\frac{1}{b}}$ for this sequence.

Things change in the case of $\sigma_n = ((\log n)(\log\log n)^b)^{-1/2}$. Here, Weber's result is

$$\log\log \left| \log \mathbb{P} \left[\sup_{n\in\mathbb{N}} |\sigma_n \xi_n| \leq \varepsilon \right] \right| \preceq \varepsilon^{-\frac{2}{b-2}},$$

for $b > 2$. We can omit this restriction and have Proposition 34, where we find the better upper bound $\varepsilon^{-\frac{2}{b}}$ for $b > 0$.

Remark 36. We observe that the upper bound of $\phi(\varepsilon)$, see (5.10), has been obtained without any use of arguments concerning the entropy of convex hulls. Consequently, Kuelbs-Li inequality (4.1) may be applied to recover that

$$\log N(\mathrm{aco}(\{\sigma_n f_n :\ n \in \mathbb{N}), \varepsilon) \preceq \varepsilon^{-2}(\log|\log\varepsilon|)^{-b}$$

as it is a result of Li and Linde (2000) obtained via the above described methods connected to majorizing measures.

5.1.2 Application: Small deviations of a Volterra type process

In this section we investigate the small deviations of a Volterra type process. Let $0 < r(b) < \exp[-2b]$ with $b > 1$ and define a kernel function

$$K_b : [0,r] \times [0,r] \to \mathbb{R},$$
$$K_b(t,s) := (t-s)^{-1/2}|\log(t-s)|^{-b}\mathbf{1}_{\{s<t\}}.$$

Although we just write r instead of $r(b)$, we ought to bear in mind that there is always dependence of b. As it is easy to confirm, we have $K_b(t,\cdot) \in L^2([0,r])$. The process $X = (X_t)_{t\in[0,r]}$ will be defined as a stochastic integral

$$X_t = \int_{[0,r]} K_b(t,s)dB_s \tag{5.28}$$

with respect to the random measure corresponding to Brownian motion $(B_s)_{s\in[0,r]}$. It is well known that the process X is centered Gaussian, see Samorodnitsky and Taqqu (1994), where an introduction on this type of processes can be found. The covariance of X is

$$\mathbb{E}X_tX_s = \int_{[0,r]} K_b(t,u)K_b(s,u)du.$$

Moreover, the process X is intimately connected to the integral operator

$$\mathcal{V}_b : L^2[0,r] \to C[0,r],\ (\mathcal{V}_b f)(t) := \int_{[0,r]} K_b(t,s)f(s)ds \tag{5.29}$$

through its dual

$$\mathcal{V}_b^* : M[0,r] \to L^2[0,r],\ (\mathcal{V}_b^*\mu)(s) = \int_{[0,r]} K_b(t,s)\mu(dt)$$

via

$$\mathbb{E}X_tX_s = \langle K_b(t,\cdot), K_s(s,\cdot)\rangle_{L^2} = \langle \mathcal{V}_b^*\delta_t, \mathcal{V}_b^*\delta_s\rangle_{L^2}.$$

As a consequence, we have

$$e_{b^2n}(K) \leq 2ae_{bn}(\mathrm{aco}(\{\mathcal{V}_b^*\delta_t : t \in [0,r]\})) \leq 2a^2 e_n(K)$$

due to Proposition 22 and hence

$$e_{b^2n}(K) \leq 2ae_{bn}(\mathrm{aco}(\{K_b(t,\cdot) : t \in [0,r]\})) \leq 2a^2 e_n(K),$$

where K denotes the unit ball of rkhs of X defined in (5.28).
To find the small deviations of X, among others we want to employ one of the Kuelbs-Li inequalities. This is why we are interested in the rate of convergence of $e_n(\mathcal{V}_b)$. A step towards this is the next lemma, which is analogous to Lemma 5 in Lifshits (2011). Its most important consequence is the relation

$$e_n(\{K_b(t,\cdot) : t \in [0,r]\}) \preceq n^{1/2-b}$$

which easily follows.

Lemma 37. *Let $b > 1$ and $0 \leq t \leq t+u \leq r$. Then we have*

$$\|K_b(t+u,\cdot) - K_b(t,\cdot)\|_2^2 \leq (1/2 + 1/(2b-1))^{1/2} |\log u|^{1/2-b}.$$

Proof. We begin to estimate

$$\begin{aligned}
&\|K_b(t+u,\cdot) - K_b(t,\cdot)\|_2^2 \\
&= \int_0^t \Big((t-s)^{-1/2}|\log(t-s)|^{-b} - (t+u-s)^{-1/2}|\log(t+u-s)|^{-b}\Big)^2 ds \\
&\quad + \int_t^{t+u} (t+u-s)^{-1}|\log(t+u-s)|^{-2b} ds \\
&= \int_0^t \Big(v^{-1/2}|\log v|^{-b} - (v+u)^{-1/2}|\log(v+u)|^{-b}\Big)^2 dv + \int_0^u v^{-1}|\log v|^{-2b} dv \\
&\overset{(a)}{\leq} \int_u^t \Big(v^{-1/2}|\log v|^{-b} - (v+u)^{-1/2}|\log(v+u)|^{-b}\Big)^2 dv + 2\int_0^u v^{-1}|\log v|^{-2b} dv \\
&\leq u^2 \int_u^t \Big(v^{-3/2}|\log v|^{-b}(-1/2 + b|\log r|^{-1})\Big)^2 dv + 2\int_0^u v^{-1}|\log v|^{-2} dv \\
&\overset{(b)}{\leq} u^2 \int_u^t v^{-3}|\log v|^{-2b} dv + 2\int_0^u v^{-1}|\log v|^{-2b} dv \\
&\leq |\log u|^{-2b} + 1/(2b-1)|\log u|^{1-2b} \\
&\leq |\log u|^{1-2b}(1/|\log r| + 1/(2b-1)).
\end{aligned}$$

Note: Estimates (a) and (b) hold in both cases $u \leq t$ and $t \leq u$. Estimate (b) holds because of mean value theorem, moreover $(-1/2 + b|\log r|^{-1})^2 \leq 1$ is valid. $\square$

The result of the next proposition is contained in Lifshits (2011), however, not its proof, which seems to be unpublished. This is why we include it; the proof of the lower bound is given in the version of Lacey (2008) communicated by W. Linde.

Proposition 38. *Let $\mathcal{V}_b$ be the integral operator defined in* (5.29) *and* $b > 1$. *Then*

$$e_n(\mathcal{V}_b) \approx n^{-1/2}(\log n)^{1-b}.$$

Proof. Let us start considering the lower bound. To prove

$$e_n(\mathcal{V}_b) \succeq n^{-1/2}(\log n)^{1-b},$$

it is sufficient to find a $cn^{-1/2}(\log n)^{1-b}$-distance net consisting of 2^n points from $\mathcal{V}_b(B_{L_2})$ with a constant $c > 0$ independent of n. We begin with setting weights

$$\rho_n^2 := \int_0^{\frac{r}{n}} \frac{1}{x|\log x|^{2b} dx} = \frac{1}{2b-1}\left|\log \frac{r}{n}\right|^{1-2b}$$

and functions

$$g_n(x) := \frac{\mathbf{1}_{[0,\frac{r}{n})}(x)}{\rho_n(\frac{r}{n} - x)^{1/2}|\log(\frac{r}{n} - x)|^b} \tag{5.30}$$

for $n \in \mathbb{N}$. Of course, these functions are constructed in such a way that $\|g_n|L^2\| = 1$. We further define

$$g_{j,n}(x) := g_n\left(x - \frac{r(j-1)}{n}\right), \ j = 1, \cdots, n. \tag{5.31}$$

For notational simplicity, let n be a square number in the following. This does not restrict generality. We define functions $f_J := |J|^{-1/2}\sum_{j\in J} g_{j,n}$ for all subsets $J \subset \{1, \cdots, n\}$. We consider the set

$$\Phi := \left\{f_J : \ J \subseteq \{1, \cdots, n\}, |J| = \sqrt{n}\right\}, \tag{5.32}$$

which will constitute a suitable distance net $\mathcal{V}_b(\Phi)$ as desired. To show this, we note that by Stirling's formula

$$\begin{aligned} |\Phi| &= \binom{n}{\sqrt{n}} \\ &\approx \frac{\sqrt{2\pi n}(n/e)^n}{\sqrt{2\pi\sqrt{n}}(\sqrt{n}/e)^{\sqrt{n}}\sqrt{2\pi(n-\sqrt{n})}((n-\sqrt{n})/e)^{n-\sqrt{n}}} \\ &\approx \frac{(\sqrt{n})^{\sqrt{n}}}{\sqrt{2\pi\sqrt{n}}(1-1/\sqrt{n})^{n-\sqrt{n}}}. \end{aligned} \tag{5.33}$$

We take the logarithm of the right hand side in (5.33) which leads to

$$\sqrt{n}\log\sqrt{n} - \log\sqrt{2\pi\sqrt{n}} - \sqrt{n}\log\left(1-\frac{1}{\sqrt{n}}\right)^{\sqrt{n}} + \sqrt{n}\log\left(1-\frac{1}{\sqrt{n}}\right)$$

and hence we see that

$$\log|\Phi| \approx \sqrt{n}\log n. \tag{5.34}$$

For $f_J \neq f_K$ let j be the least element in the symmetric difference of J and K, i.e., $j := \inf(J\cup K)\setminus(J\cap K)$.

$$\begin{aligned} \|\mathcal{V}_b(f_J) - \mathcal{V}_b(f_K)\|_\infty &\geq |\mathcal{V}_b(f_J - f_K)(\frac{jr}{n})| \\ &= \int_0^{(jr)/n} \frac{g_{j,n}(x)}{(\frac{jr}{n}-x)^{1/2}|\log(\frac{jr}{n}-x)|^b}dx \\ &= n^{-1/4}\rho_n^{-1}\int_{r(j-1)/n}^{(jr)/n} \frac{1}{(\frac{jr}{n}-x)|\log(\frac{jr}{n}-x)|^{2b}}dx \\ &= n^{-1/4}\rho_n^{-1}\int_0^{r/n}\frac{1}{x|\log x|^{2b}}dx \\ &= n^{-1/4}\rho_n \\ &= (2b-1)^{-1/2}n^{-1/4}|\log\frac{r}{n}|^{1/2-b} \\ &\geq (2b-1)^{-1/2}(\sqrt{n}\log n)^{-1/2}(\log\frac{n}{r})^{1-b}. \end{aligned}$$

Regarding (5.34), this suffices to conclude $e_k(\mathcal{V}_b) \succeq k^{-1/2}(\log k)^{1-b}$.

For the upper bound we make use of Lemma 37 which implies

$$e_n(\{K_b(t,\cdot):\ t\in[0,r]\}) \preceq n^{1/2-b}. \tag{5.35}$$

By this, the general estimate of part (iii) of Theorem 2 yields that

$$e_n(\mathcal{V}_b^*) = e_n(\mathrm{aco}(\{K_b(t,\cdot):\ t \in [0,r]\})) \preceq n^{-1/2}(\log n)^{1-b}$$

which means

$$e_n(\mathcal{V}_b) \preceq n^{-1/2}(\log n)^{1-b}$$

by Corollary 20. □

We can now formulate the main result of this section, its proof is quite short after all the preparations.

Proposition 39. *Let X be the Gaussian process defined in* (5.28)*. Its small deviations are given by*

$$\log \phi(\varepsilon) \approx \varepsilon^{-\frac{1}{b-1}}. \tag{5.36}$$

Proof. For the upper bound let us employ (5.35), which is equivalent to $\log N(\{K_b(t,\cdot):\ t \in [0,r]\},\varepsilon) \preceq \varepsilon^{-\frac{2}{2b-1}}$. From Theorem 14 it follows that

$$\log \phi(\varepsilon) \preceq \varepsilon^{-\frac{1}{b-1}}.$$

For the lower bound, Proposition 38 and Proposition 32 give

$$\log \phi(\varepsilon) \succeq \varepsilon^{-\frac{1}{b-1}},$$

which is finally the result. □

5.2 Kuelbs-Li inequalities in view of critically large entropy

We explained in the introduction, why it would be attractive to have the relation

$$\log N(K, \varepsilon/\sqrt{2\phi(\varepsilon)}) \succeq \phi(\varepsilon) \tag{5.37}$$

instead of

$$\log N(K, \varepsilon/\sqrt{2\phi(\varepsilon)}) \succeq \phi(2\varepsilon) \tag{5.38}$$

as a general result connecting the metric entropy $\log N(K,\varepsilon)$ of the unit ball of the rkHs of the Gaussian process X and the small ball function $\phi(\varepsilon) = -\log \mathbb{P}\left[\sup_{t\in T} |X_t|\right]$.

As a further motivation let us remark that – if (5.37) were fulfilled – the optimality problem of Theorem 2, part (iv) would be solved: For the proof of Proposition 9 in Aurzada and Lifshits (2008), a special Gaussian process $(X_t)_{t\in T}$ is constructed, whose small deviations have lower bound

$$\log\log\phi(\varepsilon) \succeq \varepsilon^{-\frac{2}{\beta-2}},\ \beta > 2.$$

This connected with (5.37) would yield

$$\log N(K,\varepsilon) \succeq \varepsilon^{-2}(\log|\log\varepsilon|)^{2-\beta},$$

while $\log N(T,\varepsilon) \preceq \varepsilon^{-2}|\log\varepsilon|^{-\beta}$ is satisfied.
However, relation (5.37) fails to hold in general as the next proposition states.

Proposition 40. *There is a compact set T and a Gaussian random variable X with values in $C(T)$ such that the relation*

$$\log N(K,\varepsilon/\sqrt{2\phi(\varepsilon)}) \succeq \phi(\varepsilon),$$

does not hold.

Proof. We adopt Example 3 from Aurzada and Lifshits (2008). Let $(f_k)_{k\in\mathbb{N}}$ be the canonical base of l_2, $\sigma_k = k^{-1}(\log(k+1))^{-b}$, $b > 1$ and let

$$T := \sum_{k\in\mathbb{N}}\{+\sigma_k f_k, -\sigma_k f_k\}.$$

Then we define the process X as usual as

$$X_t := \sum_{k\in\mathbb{N}}\langle t, f_k\rangle\xi_k = \sum_{k\in\mathbb{N}} t_k\sigma_k\xi_k,\ t_k = \operatorname{sgn}(\langle t, f_k\rangle),\ t\in T. \tag{5.39}$$

We have the equality

$$\sup_{t\in T}|X_t| = \sum_{k\in\mathbb{N}}\sigma_k|\xi_k|,$$

i.e., the supremum of the process equals the l_1-norm of the sequence $(\sigma_k\xi_k)_{k\in\mathbb{N}}$ and we know from Corollary 2.1 in Aurzada (2006) that

$$\log\left|\log\mathbb{P}\left[\sup_{t\in T}|X_t|\le\varepsilon\right]\right| \approx \varepsilon^{-\frac{1}{b-1}}. \tag{5.40}$$

In the next step, we want to compute the asymptotic rate of $\log N(\mathrm{aco}(T), \varepsilon)$. The problem in fact reduces to finding the entropy numbers of the corresponding diagonal operator $D_\sigma : l_\infty \to l_2$. The sequence (σ_k) fulfills $\sigma_k \approx \sigma_{2k}$ and $\sup_{n\geq k} \frac{\sigma_n}{\sigma_k} \cdot \left(\frac{n}{k}\right)^\alpha < +\infty$ for some $\alpha > \frac{1}{2}$, which are the conditions to apply Theorem 2.2 from Kühn (2005). We get

$$e_n(D_\sigma : l_\infty \to l_2) \approx n^{1/2}\sigma_n = n^{-1/2}(\log n)^{-b}, \tag{5.41}$$

so

$$\log N(\mathrm{aco}(T), \varepsilon) \approx \varepsilon^{-2}|\log\varepsilon|^{-2b}. \tag{5.42}$$

Now we suppose,

$$\log N(K, \varepsilon/\sqrt{2\phi(\varepsilon)}) \succeq \phi(\varepsilon)$$

would hold for the process defined in (5.39).
From (5.40) we know that there is a constant $c > 0$ such that

$$\phi(\varepsilon) \geq \exp[c\varepsilon^{-\frac{2}{2b-2}}]$$

for ε sufficiently small. We can compute that

$$\limsup_{\varepsilon\downarrow 0} \varepsilon^{-2}(\log\varepsilon^{-2}\phi(\varepsilon))^{2-2b} \leq \limsup_{\varepsilon\downarrow 0} \varepsilon^{-2}(\log\varepsilon^{-2}\exp[c\varepsilon^{-\frac{2}{2b-2}}])^{2-2b} < +\infty.$$

This leads to

$$\begin{aligned}
&\limsup_{\varepsilon\downarrow 0} \frac{\varepsilon^{-2}2\phi(\varepsilon)(\log\varepsilon^{-2}2\phi(\varepsilon))^{2-2b}}{\log N(K, \varepsilon/\sqrt{2\phi(\varepsilon)})} \\
&\leq \limsup_{\varepsilon\downarrow 0} \frac{\phi(\varepsilon)}{\log N(K, \varepsilon/\sqrt{2\phi(\varepsilon)})} \limsup_{\varepsilon\downarrow 0} \varepsilon^{-2}2(\log\varepsilon^{-2}2\phi(\varepsilon))^{2-2b} \\
&< +\infty.
\end{aligned}$$

Setting $\delta := \varepsilon/\sqrt{2\phi(\varepsilon)}$, we get

$$\log N(K, \delta) \succeq \delta^{-2}|\log\delta^2|^{2-2b} \approx \delta^{-2}|\log\delta|^{2-2b}. \tag{5.43}$$

If we compare (5.43) with (5.42), we find a contradiction using Proposition 21. Therefore, (5.37) cannot hold in general. $\square$

Remark 41.

(i). In the proof of Proposition 40, the special feature

$$\log N(T,\varepsilon) \approx \log N(\mathrm{aco}(T),\varepsilon)$$

was not necessary to achieve the contradiction – we could also have used sets T fulfilling $\log N(T,\varepsilon) \approx \varepsilon^{-2}|\log\varepsilon|^{-2b}$ and $\log N(\mathrm{aco}(T),\varepsilon) \approx \varepsilon^{-2}|\log\varepsilon|^{a-2b}$ for arbitrary $0 \le a < 2$. Such sets exist and may be constructed similarly to the set T in proof of Proposition 42.

(ii). Although we know that the number two on the right hand side in (2.3) cannot be omitted in general, for the example constructed in Proposition 42

$$\log N(K,\varepsilon/\sqrt{2\phi(\varepsilon)}) \succeq \phi(\varepsilon)$$

is valid. This phenomenon occurs for other quite natural examples as well.

Let us now turn to the second announced question which is whether such a one-to-one correspondence (2.5) as in the regular case still holds in the critical case. With this in mind, we first provide an upper bound for $\log N(K,\varepsilon)$ given an upper bound for $\phi(\varepsilon)$. What matters for our considerations is rather its being best possible than the special rates.

Proposition 42. *Let $b > 1$. Then $\log\phi(\varepsilon) \preceq \varepsilon^{-\frac{1}{b-1}}$ implies*

$$\log N(K,\varepsilon) \preceq \varepsilon^{-2}|\log\varepsilon|^{2-2b}.$$

This estimate is best possible.

Proof. The upper bound for $\log N(K,\varepsilon)$ can be calculated using relation (4.1).

Now let $(f_k)_{k\in\mathbb{N}}$ be an orthonormal system of l_2. Define

$$T_1 := \{(\log k)^{1/2-b} f_{2k+1} : \ k \in \mathbb{N}\}$$

and

$$T_2 := \sum_{k=1}^{\infty} \{k^{-1}(\log(k+1))^{-b} f_{2k}, -k^{-1}(\log(k+1))^{-b} f_{2k}\}.$$

It follows from the proof of Proposition 5.5 in Carl, Kyrezi, and Pajor (1999) that

$$\log N(T_1,\varepsilon) \preceq \varepsilon^{-\frac{1}{b-1/2}} \ \text{and} \ \log N(\mathrm{aco}(T_1)\varepsilon) \approx \varepsilon^{-2}|\log\varepsilon|^{2-2b}. \tag{5.44}$$

We remind (5.42) to get

$$\log N(T_2,\varepsilon) \preceq \varepsilon^{-2}|\log\varepsilon|^{-2b} \text{ and } \log N(\mathrm{aco}(T_2),\varepsilon) \approx \varepsilon^{-2}|\log\varepsilon|^{-2b}.$$

Let $(X_1(t))_{t\in T_1}$ and $(X_2(t))_{t\in T_2}$ be the corresponding Gaussian processes defined by (3.19).
For $\phi_i(\varepsilon) := -\log\mathbb{P}\left[\sup_{t\in T_i}|X_t| \le \varepsilon\right]$, $i=1,2$, it is known that

$$\log\phi_1(\varepsilon) \preceq \varepsilon^{-\frac{1}{b-1}} \text{ and } \log\phi_2(\varepsilon) \approx \varepsilon^{-\frac{1}{b-1}}.$$

For the first, use (5.44) in connection with Theorem 3 in Aurzada and Lifshits (2008), for the second we refer to the proof of Proposition 40. Moreover, define $T := T_1 + T_2$ and observe that

$$\log N(T,\varepsilon) \preceq \varepsilon^{-2}|\log\varepsilon|^{-2b} \text{and } \log N(\mathrm{aco}(T),\varepsilon) \approx \varepsilon^{-2}|\log\varepsilon|^{2-2b}.$$

Now it suffices to show that $\phi(\varepsilon) = -\log\mathbb{P}\left[\sup_{t\in T}|X(t)| \le \varepsilon\right] \preceq \varepsilon^{-\frac{1}{b-1}}$, where X is the Gaussian process connected to T and defined by (3.19). Note that for $t = t_1 + t_2$, $t_1 \in T_1, t_2 \in T_2$, we have $X(t) = X_1(t_1) + X_2(t_2)$ This yields

$$\sup_{t\in T}|X(t)| \le \sup_{t_1\in T_1}|X_1(t_1)| + \sup_{t_2\in T_2}|X_2(t_2)|.$$

For $\varepsilon > 0$ it follows that

$$\begin{aligned}
\mathbb{P}\left[\sup_{t\in T}|X(t)| \le \varepsilon\right] &\ge \mathbb{P}\left[\sup_{t\in T_1}|X_1(t)| + \sup_{t\in T_2}|X_2(t)| \le \varepsilon\right] \\
&\ge \mathbb{P}\left[\sup_{t\in T_1}|X_1(t)| \le \frac{\varepsilon}{2},\ \sup_{t\in T_2}|X_2(t)| \le \frac{\varepsilon}{2}\right] \\
&= \mathbb{P}\left[\sup_{t\in T_1}|X_1(t)| \le \frac{\varepsilon}{2}\right]\mathbb{P}\left[\sup_{t\in T_2}|X_2(t)| \le \frac{\varepsilon}{2}\right],
\end{aligned}$$

where we use orthogonality of the sets T_1 and T_2 in Hilbert space for the last step. Hence,

$$\phi(\varepsilon) \le \phi_1(\frac{\varepsilon}{2}) + \phi_2(\frac{\varepsilon}{2})$$

and therefore

$$\log\phi(\varepsilon) \preceq \varepsilon^{-\frac{1}{b-1}}. \tag{5.45}$$

Thus, we have shown that there is a Gaussian process with continuous sample paths a.s., whose small ball function satisfies (5.45) and whose unit ball of rkHs has metric entropy

$$\log N(K,\varepsilon) \approx \varepsilon^{-2}|\log\varepsilon|^{2-2b}.$$

□

Let us relate Proposition 42 to the preceding Proposition 40 and realize that we cannot expect a general inverse assertion in the way that $\log \phi(\varepsilon) \succeq \varepsilon^{-\frac{1}{b-1}}$ would imply $\log N(K,\varepsilon) \succeq \varepsilon^{-2}|\log\varepsilon|^{2-2b}$.
For this, we recall ourselves of the process $(X_t)_{t\in T}$ defined in (5.39) which possesses the properties $\log\phi(\varepsilon) \approx \varepsilon^{-\frac{1}{b-1}}$ and $\log N(\mathrm{aco}(T)) \approx \varepsilon^{-2}|\log\varepsilon|^{-2b}$.
We concentrate our observation in the next corollary.

Corollary 43. *There are Gaussian processes* $X_1 = (X_t)_{t\in T_1},\ X_2 = (X_t)_{t\in T_2}$ *such that the corresponding small ball functions* ϕ_1, ϕ_2 *and the corresponding unit balls of rkHs* K_1, K_2 *fulfill the relations*

$$\log\phi_1(\varepsilon) \approx \log\phi_2(\varepsilon) \approx \varepsilon^{-\frac{1}{b-1}},$$

$$\log N(K_1,\varepsilon) \approx \varepsilon^{-2}|\log\varepsilon|^{-2b} \text{ and } \log N(K_2,\varepsilon) \approx \varepsilon^{-2}|\log\varepsilon|^{2-2b}.$$

In the same way as the rate of $\log\phi(\varepsilon)$ does not determine the rate of $\log N(K,\varepsilon)$, we find that the rate of $\log N(K,\varepsilon)$ does not determine the rate of $\log\phi(\varepsilon)$ as can be seen from the next statement.

Corollary 44. *There are Gaussian processes* $X_1 = (X_t)_{t\in T_1},\ X_2 = (X_t)_{t\in T_2}$ *such that the corresponding small ball functions* ϕ_1, ϕ_2 *and the corresponding unit balls of rkHs* K_1, K_2 *fulfill the relations*

$$\log N(K_1,\varepsilon) \approx \log N(K_2,\varepsilon) \approx \varepsilon^{-2}|\log\varepsilon|^{-2b},$$

$$\log\phi_1(\varepsilon) \approx \log\varepsilon^{-\frac{1}{b-1}}, \text{ and } \log\phi_2(\varepsilon) \preceq \varepsilon^{-1/b}.$$

Proof. Again, let X_1 be the process defined in (5.39). Define X_2 with $b > 1$ via (3.19) using the set $T_2 := \{(\log k + 1)^{-\frac{1+2b}{2}} f_k :\ k \in \mathbb{N}\} \subseteq l_2$, and from proof of Proposition 5.5 in Carl, Kyrezi, and Pajor (1999) it follows that

$$\log N(T_2,\varepsilon) \preceq \varepsilon^{-\frac{1}{b+1/2}} \text{ and } \log N(\mathrm{aco}(T_2)) \approx \varepsilon^{-2}|\log\varepsilon|^{-2b},$$

while we have $\log\phi_2(\varepsilon) \preceq \varepsilon^{-1/b}$ by Theorem 14. □

It would be interesting to know whether there are Gaussian processes $X_1 = (X_t)_{t\in T_1}, X_2 = (X_t)_{t\in T_2}$ with sample paths in $C(T)$ a.s. such that

$$\log N(T_1,\varepsilon) \approx \log N(T_2,\varepsilon) \textit{ and } \log N(\mathrm{aco}(T_1),\varepsilon) \approx \log N(\mathrm{aco}(T_2),\varepsilon), \tag{5.46}$$

while $\log\phi_1 \not\approx \log\phi_2$.
Note that for the examples related to the preceding corollaries either the first or the second condition of (5.46) is violated.

5.3 Variations on the Gao set

As remarked in the introductory chapter, Gao (2001) was the first to prove estimate (vi) of Theorem 2 for the special case of $b = 0$, onward he constructed a set A with

$$e_n(A) \preceq n^{-1/2} \quad \text{and} \quad e_n(\mathrm{aco}(A)) \succeq n^{-1/2} \log n$$

proving the sharpness of his upper estimate. Creutzig and Steinwart (2002) extended his ideas to the case of p-Banach spaces for $1 < p \leq 2$, hence Theorem 2 part (vi) is included there. Since estimates (iv) and (v) of Theorem 2 are not known to be best possible, it is self-evident to consider the Gao set A whose construction can be modified to fit in the setting of $b \geq 2$ in the sense that $\log N(A, \varepsilon) \preceq \varepsilon^{-2} |\log \varepsilon|^{-b}$. As to a lower bound for $\log N(\mathrm{aco}(A), \varepsilon)$, we will provide some observations after having introduced the set A.

From Aurzada and Lifshits (2008), Proposition 9, we borrow a construction which was first for a Gaussian process. It rather resembles the original construction in Gao (2001) than the construction in Creutzig and Steinwart (2002).

Let $(f_k)_{k\in\mathbb{N}}$ be the canonical orthonormal base of l_2 and $(N_k)_{k\in\mathbb{N}}$ a partition of $\mathbb{N}$ in the way that $d_k := |N_k| = 2^{2^{2k}}$.
For all $k \in \mathbb{N}$ and $b \in \mathbb{R}$ we define sets

$$D_k := \{\pm 2^{-k} k^{-b} f_i : \; i \in N_k\}$$

and with these a set A as the Minkowski sum

$$A := \sum_{k\in\mathbb{N}} D_k. \tag{5.47}$$

We keep in mind that all D_k as well as A depend on b, which will not be reflected in the designation for notational simplicity.

The next lemma provides an upper bound for the entropy of A. We will give the proof since A differs from both sets constructed in Gao (2001) and Creutzig and Steinwart (2002), respectively, although it works with the same ideas.

Lemma 45. *Let A be as above and $b \in \mathbb{R}$. Then we have*

$$e_n(A) \preceq n^{-1/2} (\log n)^{-b}.$$

Proof. As a preliminary consideration we estimate the cardinality of the set $\sum_{k=1}^{N} D_k$ by

$$\left|\sum_{k=1}^{N} D_k\right| \leq \prod_{k=1}^{N} |D_k| \leq \prod_{k=1}^{N} 2d_k \leq 2^{N+\sum_{k=1}^{N} 2^{2k}} \leq d_{N+2}.$$

By $c_0 > 0$, we denote a constant for which

$$\sum_{k=M}^{\infty} 2^{-k} k^{-b} \leq c_0 2^{-M} M^{-b}$$

is valid for all $M \in \mathbb{N}$.

Now let $n > 2^8$ and fix an $N \in \mathbb{N}$ so that $d_{N+3} \leq 2^{n-1}$. If we choose N to be maximal under this condition, we have

$$d_{N+3} \leq 2^{n-1} \leq d_{N+4.} \tag{5.48}$$

The set A will be divided into the two parts

$$A_1 := \sum_{k=1}^{N-1} D_k \quad \text{and } A_2 := \sum_{k=N}^{\infty} D_k,$$

with each of them being dealt with separately.

Let $\varepsilon := c_0 2^{-N} N^{-b}$. We can afford to spend one $\varepsilon-$ball for each point in A_1, since their number is bounded by

$$|A_1| \leq \left|\sum_{k=1}^{N-1} D_k\right| \leq d_{N+1}.$$

The set A_2 is small, i.e.

$$\|A_2\| \leq \sum_{k=N}^{\infty} \|D_k\| \leq \varepsilon.$$

Hence, only one $\varepsilon-$ball is needed to cover A_2.

For the entropy numbers we then have

$$\begin{aligned} e_{2n}(A) = e_{2n}(A_1 + A_2) &\leq e_n(A_1) + e_n(A_2) \\ &\leq c_1\varepsilon \leq c_2 2^{-N} N^{-b} \leq c_3 n^{-1/2} (\log n)^{-b}, \end{aligned}$$

wherefrom we get the assertion with regard to (5.48). $\square$

The following lemma is due to Gao (2001), we quote it in the formulation of Creutzig and Steinwart (2002). It is a key step for the lower estimate, at least in the case of $b < 1$.

Lemma 46 (Gao)**.** *Let $H_1, \cdots, H_N$ be Hilbert spaces. We equip the product space $H_1 \times \cdots \times H_N$ with the product norm*

$$\|(x_1, \cdots, x_N)\|_2 := \left(\sum_{i=1}^{N} \|x_i\|^2 \right)^{1/2}.$$

Then for all subsets $A_i \subseteq H_i$ and every $n \geq 6$ we have

$$N^{1/2} \min_{i \leq N} e_{n+1}(A_i) \leq 4 e_{\lfloor \frac{nN}{3} \rfloor}(A_1 \times \cdots \times A_N).$$

We need another lemma.

Lemma 47. *Let $A, B \subseteq E$ two symmetric subsets of a linear space E, i.e. $A = -A$ and $B = -B$. Then we have* $\mathrm{aco}(A + B) = \mathrm{aco}(A) + \mathrm{aco}(B)$.

Proof. The inclusion $\mathrm{aco}(A + B) \subseteq \mathrm{aco}(A) + \mathrm{aco}(B)$ is obvious. For the other one take

$$x = \sum_{i=1}^{n} \lambda_i a_i + \sum_{j^1}^{m} \mu_j b_j \in \mathrm{aco}(A) + \mathrm{aco}(B); \ \sum_{j=1}^{m} |\mu_j| = \sum_{i=1}^{n} |\lambda_i| = 1.$$

Since A, B are symmetric we may assume w.l.o.g. that all λ_i, μ_j are non-negative. Then we can rewrite x as

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \lambda_i \mu_j (a_i + b_j) = \sum_{i=1}^{n} \lambda_i a_i + \sum_{j=1}^{m} \mu_j b_j = x.$$

Additionally, we also have $\sum_{i=1}^{n} \sum_{j=1}^{m} |\lambda_i \mu_j| = 1$, so we are done. □

Analogously to Creutzig and Steinwart (2002) we can now provide a lower bound for $\mathrm{aco}(A)$ in terms of entropy numbers.

Lemma 48. *Let A be as in* (5.47) *and $b \in \mathbb{R}$. Then we have*

$$e_n(\mathrm{aco}(A)) \succeq n^{-1/2} (\log n)^{1-b}.$$

Proof. Let $n \geq 32$ and fix an $N \in \mathbb{N}$ so that

$$(N-1)\log_2 d_{2(N-1)} \leq 3n \leq N \log_2 d_{2N}. \tag{5.49}$$

For short we write $m := \log_2 d_{2N}$ and observe $n \leq Nm/3$, which implies

$$e_n(\mathrm{aco}(A)) \geq e_{\lfloor Nm/3 \rfloor}(\mathrm{aco}(\sum_{k=1}^{\infty} D_k)) \geq e_{\lfloor Nm/3 \rfloor}(\sum_{k=N+1}^{2N} \mathrm{aco}(D_k))$$

while keeping Lemma 47 in mind. Furthermore, we have

$$\begin{aligned} e_{\lfloor Nm/3 \rfloor}(\sum_{k=N}^{2N} \mathrm{aco}(D_k)) &\geq e_{\lfloor Nm/3 \rfloor}(\mathrm{aco}(D_{N+1}) \times \cdots \times \mathrm{aco}(D_{2N})) \\ &\geq \frac{1}{4} N^{1/2} \min_{N+1 \leq k \leq 2N} e_{m+1}(\mathrm{aco}(D_k)) \end{aligned} \tag{5.50}$$

applying Gao's lemma, Lemma 46, in the last step.

Note that
$$\log_2 d_k \leq m \leq d_k^{1/2}.$$
Hence, using a well known theorem of Schütt (1984), we can go on estimating

$$\begin{aligned} e_{m+1}(\mathrm{aco}(D_k)) &= 2^{-k} k^{-b} e_{m+1}(\mathrm{aco}(\{\pm f_i : \ i \in N_k\})) \\ &= 2^{-k} k^{-b} e_{m+1}(id : l_1^{d_k} \to l_2^{d_k}) \\ &\geq c_1 2^{-k} k^{-b} \left(\frac{\log_2 \frac{d_k}{m+1}}{m+1} \right)^{1/2} \\ &\geq c_2 2^{-k} k^{-b} \left(\frac{\log_2 d_k^{1/2}}{m+1} \right)^{1/2} \\ &\geq c_3 k^{-b} m^{-1/2} \end{aligned}$$

and we can proceed with (5.50) leading to

$$e_n(\mathrm{aco}(A)) \geq c_4 N^{1/2} N^{-b} m^{-1/2} = c_5 N^{-1/2} N^{1-b} m^{-1/2} = c_5 (mN)^{-1/2} N^{1-b}.$$

Remembering (5.49) this finally gives

$$e_n(\mathrm{aco}(A)) \geq c_6 n^{-1/2} (\log_2 n)^{1-b}, \tag{5.51}$$

where the constant depends on the choice of $b \in \mathbb{R}$. □

In fact, Lemma 48 shows the sharpness of Theorem 2 part (vi), but is far away from showing sharpness of estimates (iv) and (v).

Now the question is, whether the multiplicative increase of $\log n$ in the lower bound is intrinsic to the construction of this set, so to speak one could suppose that

$$e_n(\mathrm{aco}(A)) \approx n^{-1/2}(\log n)^{1-b}$$

for all $b \in \mathbb{R}$.

As an alternative for $b \geq 1$, $e_n(\mathrm{aco}(A))$ could have greater entropy reflecting that the Gao lemma does not perform well in this case.

In Linde (2008), a general lower bound for the entropy numbers of an operator $v : H \to C(T)$ is developed, where H denotes a Hilbert space and T a compact metric space. Certain numbers $\tau_n(u), n \in \mathbb{N}$, are defined, originating in the theory of local non-determinism of stochastic processes. They are rewritten as

$$\tau_n(v) = \sup\{\min_{1\leq j\leq n} dist(\{g_j\}, span\{g_1, \cdots, g_{j-1}\})|g_1, \cdots, g_n \in A_v\} \quad (5.52)$$

see Corollary 2 loc. cit. , where $A_v := \{v^*\delta_t | t \in T\}$. With these, the following Theorem is formulated.

Theorem 49 (Linde)**.** *Let H be a Hilbert space and T be a compact metric space. For all $v : H \to C(T)$, we have*

$$2e_n(v) \geq n^{-1/2}\tau_n(v), \quad n = 1, 2, \cdots .$$

Now we can state a lower bound for the entropy of the absolutely convex hull of A.

Proposition 50. *Let A be as in* (5.47) *and $b > 1$. Then we have*

$$e_n(\mathrm{aco}(A)) \succeq n^{-1/2}(\log n)^{-1/2}(\log\log n)^{-b}. \quad (5.53)$$

Proof. We take the closure $\overline{A}$ of the set constructed in (5.47) and define an operator $u : l_1(\overline{A}) \to l_2$ in the same way as in (3.18) .
Its dual $u^* : l_2 \to l_\infty(\overline{A})$ in fact is a mapping into $C(\overline{A})$ since it is given by $(u^*h)(a) = \langle a, h\rangle$ for all $h \in l_2$ and all $a \in \overline{A}$.
The dual of the operator $u^* : l_2(\overline{A}) \to C(\overline{A})$ is the mapping $u^{**} : M(\overline{A}) \to l_2$ given by

$$\langle h, u^{**}\mu\rangle = \int_{\overline{A}} u^*(h)d\mu, \quad \mu \in M(\overline{A}),$$

hence for a point measure $\delta_a,\ a \in \overline{A}$, we have $u^{**}(\delta_a) = a$ by means of Riesz representation.

In view of Theorem 49 it suffices to find a lower bound for the numbers $\tau_n(u^*)$ to get a lower bound for $e_n(u^*)$, which will run into a lower bound for $e_n(\mathrm{aco}(A))$ by Proposition 21.

We have the simple lower estimate

$$\tau_n(u^*) \geq \sup\{\min_{1\leq j\leq n} dist(\{g_j\}, span\{g_1, \cdots, g_{j-1}\})|g_1, \cdots, g_n \in A\} \quad (5.54)$$

Let us try to evaluate the right hand side in (5.54). For this, note that each g_i may be written as $g_i = \sum_{k=1}^{\infty} \sigma_k t_k^i f_k^i$ with $t_k^i \in \{+1, -1\}$ and $f_k^i \in \{f_l |\ l \in N_k\}$. Then

$$\begin{aligned} dist^2(\{g_j\}, span\{g_1, ..., g_{j-1}\}) &= \inf_{\lambda_i \in \mathbb{R}} \|g_j - \sum_{i=1}^{j-1} \lambda_i g_i\|^2 \\ &= \inf_{\lambda_i} \|\sum_{k=1}^{\infty} \sigma_k t_k^j f_k^j - \sum_{i=1}^{j-1} \lambda_i \sum_{k=1}^{\infty} \sigma_k t_k^i \lambda_i f_k^i\|^2 \\ &= \inf_{\lambda_i} \sum_{k=1}^{\infty} \sigma_k^2 \|t_k^j f_k^j - \sum_{i=1}^{j-1} t_k^i \lambda_i f_k^i\|^2 \\ &= \inf_{\lambda_i} \sum_{k=1}^{\infty} \sigma_k^2 \|f_k^j - \sum_{i=1}^{j-1} \lambda_i f_k^i\|^2 \end{aligned}$$

Let us choose special $g_i's$. For $k \geq \left\lceil \frac{1}{2} \log_2 \log_2 n \right\rceil =: n_0$ we have the freedom to choose all $f_k^i,\ 1 \leq i \leq n$ to be pairwise distinct since then $|N_k| \geq n$. We do not impose restrictions on the choice of all t_k^i as well as on the choice of all f_k^i for $k < n_0$. Then

$$\tau_n^2(u) \geq \min_{1\leq j\leq n} \inf_{\lambda_i} \sum_{k=1}^{\infty} \sigma_k^2 \|f_k^j - \sum_{i=1}^{j-1} \lambda_i f_k^i\|^2 \geq \sum_{k=n_0}^{\infty} \sigma_k^2 \quad (5.55)$$

Now we only have to do a calculation exercise.

$$
\begin{aligned}
&\sum_{k=n_0}^{\infty} \sigma_k^2 \geq (\log 2)^{2b} \int_{n_0}^{\infty} (2^x)^{-2} (\log 2^x)^{-2b} dx = \left[y := 2^x, dx = \frac{dy}{\log 2\ y} \right] \\
&= (\log 2)^{2b+1} \int_{2^{n_0}}^{\infty} y^{-3} (\log y)^{-2b} dy \\
&= (\log 2)^{2b} \left(\left[-\frac{1}{2} y^{-2} (\log y)^{-2b} \right]_{2^{n_0}}^{\infty} - b \int_{2^{n_0}}^{\infty} y^{-3} (\log y)^{-2b-1} dy \right) \\
&= (\log 2)^{2b+1} \left(\frac{1}{2(\log 2)^{2b}} (2^{n_0})^{-2} n_0^{-2b} - b \left[-\frac{1}{2} y^{-2} (\log y)^{-2b-1} \right]_{2^{n_0}}^{\infty} \right) \\
&\quad + (\log 2)^{2b+1} \frac{2b+1}{2} \int_{2^{n_0}}^{\infty} y^{-3} (\log y)^{-2b-2} \\
&\geq \frac{1}{2} (2^{n_0})^{-2} n_0^{-2b} - \frac{b}{2 \log 2} (2^{n_0})^{-2} n_0^{-2b-1} \\
&= (2^{n_0})^{-2} n_0^{-2b} \left(\frac{\log 2}{2} - \frac{b}{2} n_0^{-1} \right) \overset{(a)}{\geq} \frac{1}{4} (2^{n_0})^{-2} n_0^{-2b} \\
&\overset{(b)}{\geq} \frac{1}{16} \left(2^{\log_2 (\log_2 n)^{\frac{1}{2}}} \right)^{-2} (\log_2 \log_2 n + 1)^{-2b} \\
&\geq \frac{1}{16} (\log_2 n)^{-1} (\log_2 \log_2 n + 1)^{-2b},
\end{aligned}
$$

where estimate (a) is valid whenever $n \geq 2^{2^{20b}}$; for estimate (b) we take into account that $n_0 \leq \frac{1}{2} \log_2 \log_2 n + 1$. $\square$

Remark 51. Let us rewrite the assertion of Proposition 50 as

$$e_n(acoA) \succeq n^{-1/2} (\log n)^{-b} (\log n)^{b-1/2} (\log \log n)^{-b}.$$

As we see, there is a multiplicative increase of $(\log n)^{b-1/2} (\log \log n)^{-b}$, which – for $b > 3/2$ – is a better estimate than just $\log n$ as provided by the Gao lemma.

BIBLIOGRAPHY

Artstein, S., V. Milman, and S. J. Szarek (2004). “Duality of metric entropy.” English. In: *Ann. Math. (2)* 159, pp. 1313–1328. DOI: 10.4007/annals.2004.159.1313.

Aurzada, F. (2006). *Small deviation probabilities of some stochastic processes.* English. Jena: Univ. Jena, Fakultät für Mathematik und Informatik (Diss.). 74 p.

— (2008). “A short note on small deviations of sequences of i.i.d. random variables with exponentially decreasing weights.” English. In: *Stat. Probab. Lett.* 78.15, pp. 2300–2307. DOI: 10.1016/j.spl.2008.02.007.

— (2010). Private communication.

Aurzada, F. and M.A. Lifshits (2008). “Small deviation probability via chaining.” English. In: *Stochastic Processes Appl.* 118.12, pp. 2344–2368. DOI: 10.1016/j.spa.2008.01.005.

Aurzada, F. et al. (2009). “Small deviations of smooth stationary Gaussian processes.” English. Russian original. In: *Theory Probab. Appl.* 53.4, pp. 697–707. DOI: 10.1137/S0040585X97983912.

Ball, K. and A. Pajor (1991). *The entropy of convex bodies with "few" extreme points.* English. Geometry of Banach spaces, Proc. Conf., Strobl/Austria 1989, Lond. Math. Soc. Lect. Note Ser. 158, 25-32 (1991).

Bauer, H. (1990). *Maß- und Integrationstheorie. (Measure and integration theory).* German. Berlin etc.: Walter de Gruyter. xviii, 259 S.

Bogachev, V.I. (1998). *Gaussian measures. Transl. from the Russian by the author.* English. Mathematical Surveys and Monographs. 62. Providence, RI: American Mathematical Society (AMS). xii, 433 p.

Borovkov, A.A. and P.S. Ruzankin (2008). “On small deviations of series of weighted random variables.” English. In: *J. Theor. Probab.* 21.3, pp. 628–649. DOI: 10.1007/s10959-007-0130-x.

Bourbaki, N. (1987). *Elements of mathematics. Topological vector spaces. Chapters 1-5. Transl. from the French by H. G. Eggleston and S. Madan.* English. Berlin etc.: Springer-Verlag. VII, 364 p.

Bühler, B. (1999). “Finite majorizing measures.” PhD thesis. Jena: Univ. Jena, Fakultät für Mathematik und Informatik, 78 p.

Bühler, B., W.V. Li, and W. Linde (2001). *Localization of majorizing measures.* English. Balakrishnan, N. (ed.) et al., Asymptotic methods in probability and statistics with applications. Papers from the international conference, St. Petersburg, Russia, June 24-28, 1998. Boston, MA: Birkhäuser. 81-100 (2001).

Carl, B. (1982). "On a characterization of operators from $l_q$into a Banach space of type p with some applications to eigenvalue problems." English. In: *J. Funct. Anal.* 48, pp. 394–407. DOI: `10.1016/0022-1236(82)90093-3`.

— (1997). "Metric entropy of convex hulls in Hilbert spaces." English. In: *Bull. Lond. Math. Soc.* 29.4, pp. 452–458. DOI: `10.1112/S0024609397003044`.

Carl, B. and D. E. Edmunds (2003). "Gelfand numbers and metric entropy of convex hulls in Hilbert spaces." English. In: *Stud. Math.* 159.3, pp. 391–402. DOI: `10.4064/sm159-3-4`.

Carl, B., A. Hinrichs, and A. Pajor (2010). "Gelfand numbers and metric entropy of convex hulls in Hilbert spaces." English. To appear.

Carl, B., I. Kyrezi, and A. Pajor (1999). "Metric entropy of convex hulls in Banach spaces." English. In: *J. Lond. Math. Soc., II. Ser.* 60.3, pp. 871–896. DOI: `10.1112/S0024610799008005`.

Carl, B. and I. Stephani (1990). *Entropy, compactness and the approximation of operators.* English. Cambridge Tracts in Mathematics, 98. Cambridge etc.: Cambridge University Press. x, 277 p.

Cohn, D. L. (1980). *Measure theory.* English. Boston, Basel, Stuttgart: Birkhäuser. IX, 373 p.

Creutzig, J. and I. Steinwart (2002). "Metric entropy of convex hulls in type p spaces–the critical case." English. In: *Proc. Am. Math. Soc.* 130.3, pp. 733–743. DOI: `10.1090/S0002-9939-01-06256-6`.

Dudley, R. M. (1967). "The sizes of compact subsets of Hilbert space and continuity of Gaussian processes." English. In: *J. Funct. Anal.* 1, pp. 290–330. DOI: `10.1016/0022-1236(67)90017-1`.

— (1987). "Universal Donsker classes and metric entropy." English. In: *Ann. Probab.* 15, pp. 1306–1326. DOI: `10.1214/aop/1176991978`.

Dunker, T. (1998). "Small ball estimates for the fractional Brownian sheet." English. PhD thesis. Jena: Univ. Jena, Fakultät Mathematik/Informatik, 52 p.

Edmunds, D. E. and H. Triebel (1996). *Function spaces, entropy numbers, differential operators.* English. Cambridge Tracts in Mathematics. 120. Cambridge: Cambridge Univ. Press. xi, 252 p.

Fernique, X. (1997). *Gaussian random functions, Gaussian random vectors. (Fonctions aléatoires gaussiennes, vecteurs aléatoires gaussiens.)* French. Montréal: Les Publications CRM. 217 p.

Gao, F. (2001). "Metric entropy of convex hulls." English. In: *Isr. J. Math.* 123, pp. 359–364. DOI: 10.1007/BF02784136.

— (2004). "Entropy of absolute convex hulls in Hilbert spaces." English. In: *Bull. Lond. Math. Soc.* 36.4, pp. 460–468. DOI: 10.1112/S0024609304003121.

Kolmogorov, A.N. and V.M. Tikhomirov (1959). "ε-entropy and ε-capacity of sets in function spaces." Russian. In:

Kuelbs, J. (1976). "A strong convergence theorem for Banach space valued random variables." English. In: *Ann. Probab.* 4, pp. 744–771. DOI: 10.1214/aop/1176995982.

Kuelbs, J. and W. V. Li (1993). "Metric entropy and the small ball problem for Gaussian measures." English. In: *J. Funct. Anal.* 116.1, pp. 133–157. DOI: 10.1006/jfan.1993.1107.

Kühn, T. (2005). "Entropy numbers of general diagonal operators." English. In: *Rev. Mat. Complut.* 18.2, pp. 479–491.

Kwapień, S. and W. A. Woyczyński (1992). *Random series and stochastic integrals: single and multiple.* English. Probability and Its Applications. Boston: Birkhäuser. xvi, 360 p.

Lacey, M. (2008). Unpublished manuscript.

Ledoux, M. (1996). *Isoperimetry and Gaussian analysis.* English. Dobrushin, R. (ed.) et al., Lectures on probability theory and statistics. Ecole d'été de probabilités de Saint-Flour XXIV – 1994. Lectures given at the summer school in Saint-Flour, France, July 7–23, 1994. Berlin: Springer. Lect. Notes Math. 1648, 165-294 (1996).

Ledoux, M. and M. Talagrand (1991). *Probability in Banach spaces. Isoperimetry and processes.* English. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge, 23. Berlin etc.: Springer-Verlag. xii, 480 p.

Li, W. V. and W. Linde (1999). "Approximation, metric entropy and small ball estimates for Gaussian measures." English. In: *Ann. Probab.* 27.3, pp. 1556–1578. DOI: 10.1214/aop/1022677459.

— (2000). "Metric entropy of convex hulls in Hilbert spaces." English. In: *Stud. Math.* 139.1, pp. 29–45.

Li, W.V. and Q.-M. Shao (2001). *Gaussian processes: Inequalities, small ball probabilities and applications.* English. Shanbhag, D. N. (ed.) et al., Stochastic processes: Theory and methods. Amsterdam: North-Holland/Elsevier. Handb. Stat. 19, 533-597 (2001).

Lifshits, M. A. (1999). *Asymptotic behavior of small ball probabilities.* English. Grigelionis, B. (ed.) et al., Probability theory and mathematical statistics. Proceedings of the 7th international Vilnius conference, Vilnius, Lithuania, August, 12-18, 1998. Vilnius: TEV. 453-468 (1999).

— (2009). "A new proof of Gao's estimate". Unpublished manuscript.

Lifshits, M.A. (2011). "Bounds for Entropy Numbers of Some Critical Operators". In: *arXiv:1002.1377v1 [math.FA].*

Linde, W. (2008). "Nondeterminism of linear operators and lower entropy estimates." English. In: *J. Fourier Anal. Appl.* 14.4, pp. 568–587. DOI: 10.1007/s00041-008-9023-3.

Pisier, G. (1981). "Remarques sur un résultat non publie de B. Maurey." French. In: *Semin. Anal. Fonct. 1980-1981* Exp. No. V, 13 pp.

— (1989). *The volume of convex bodies and Banach space geometry.* English. Cambridge Tracts in Mathematics, 94. Cambridge etc.: Cambridge University Press. xv, 250 p.

Roy, N. M. (1987). "Extreme points of convex sets in infinite dimensional spaces." English. In: *Am. Math. Mon.* 94, pp. 409–422. DOI: 10.2307/2322725.

Rudin, W. (1973). *Functional analysis.* English. McGraw-Hill Series in Higher Mathematics. New York etc.: McGraw-Hill Book Comp. XIII, 397 p.

Samorodnitsky, G. and M. S. Taqqu (1994). *Stable non-Gaussian random processes: stochastic models with infinite variance.* English. Stochastic Modeling. New York, NY: Chapman & Hall. xviii, 632 p.

Schütt, C. (1984). "Entropy numbers of diagonal operators between symmetric Banach spaces." English. In: *J. Approximation Theory* 40, pp. 121–128. DOI: 10.1016/0021-9045(84)90021-2.

Steinwart, I. (2000). "Entropy of $C(K)$-valued operators and some applications." English. PhD thesis. Jena: Univ. Jena, Fakultät für Mathematik und Informatik, 74 p.

— (2004). "Entropy of convex hulls – some Lorentz norm results." English. In: *J. Approximation Theory* 128.1, pp. 42–52. DOI: 10.1016/j.jat.2004.04.001.

Sudakov, V.N. (1969). "Gaussian measures, Cauchy measures and ε-entropy." English. Russian original. In: *Sov. Math., Dokl.* 10, pp. 310–313.

Talagrand, M. (1993). "New Gaussian estimates for enlarged balls." English. In: *Geom. Funct. Anal.* 3.5, pp. 502–526. DOI: 10.1007/BF01896240.

— (1996). "Majorizing measures: The generic chaining." English. In: *Ann. Probab.* 24.3, pp. 1049–1103. DOI: 10.1214/aop/1065725175.

Trenkmann, K. (2009). "The Reproducing Kernel Hilbert Space of the Fractional Brownian Motion". MA thesis. Friedrich-Schiller-Universiät Jena.

Vakhaniya, N.N., V.I. Tarieladze, and S.A. Chobanyan (1987). *Probability distributions on Banach spaces. Transl. from the Russian by Wojbor A. Woyczynski.* English. Mathematics and Its Applications (Soviet Series), 14. Dordrecht etc.: D. Reidel Publishing Company. xxvi, 482 p.

Weber, M. (2010). "On small deviation of Gaussian Processes using Majorizing Measures". URL: `http://arxiv.org/abs/1012.3614`.

Ehrenwörtliche Erklärung

Hiermit erkläre ich,

- dass mir die Promotionsordnung der Fakultät bekannt ist,
- dass ich die Dissertation selbst angefertigt habe, keine Textabschnitte oder Ergebnisse eines Dritten oder eigene Prüfungsarbeiten ohne Kennzeichung übernommen habe und alle von mir benutzten Hilfsmittel, persönliche Mitteilungen und Quellen in meiner Arbeit angegeben habe,
- dass ich die Hilfe eines Promotionsberaters nicht in Anspruch genommen habe und dass Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen,
- dass ich die Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht habe.

Ich habe die gleiche, eine in wesentlichen Teilen ähnliche bzw. eine andere Abhandlung noch nicht bei einer anderen Hochschule als Dissertation eingereicht.

Jena, den ....

Unterschrift